From DEPARTMENT OF MEDICAL EPIDEMIOLOGY AND BIOSTATISTICS
Karolinska Institutet, Stockholm, Sweden

TRANSMITTED AND ACQUIRED DETERMINANTS FOR CHILDHOOD ASTHMA – FROM GENES TO TEENS

Vilhelmina Ullemar

Karolinska Institutet

Stockholm 2016

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by E-Print AB, 2016
Cover illustration by the author.
Plate layout figure on page 30 provided by Robert Karlsson, with permission.
All other figures and illustrations courtesy of the author.

ISBN 978-91-7676-400-8

Transmitted and acquired determinants for childhood asthma – from genes to teens

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Vilhelmina Ullemar

Principal Supervisor:

Professor Catarina Almqvist Malmros
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Co-supervisor(s):

Professor Paul Lichtenstein
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Associate Professor Patrik Magnusson
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Professor Göran Pershagen
Karolinska Institutet
Institute of Environmental Medicine

Opponent:

Professor John Henderson
University of Bristol
School of Social and Community Medicine

Examination Board:

Associate Professor Jette Möller
Karolinska Institutet
Department of Public Health Sciences

Associate Professor Lennart Nilsson
University of Linköping
Faculty of Medicine and Health Sciences

Professor Tomas Ekström
Karolinska Institutet
Department of Clinical Neuroscience

To my maternal grandfather Leo Ullemar, as a tribute to his legacy of academia and adventure.

ABSTRACT

The interplay between genetic and environmental factors is central to childhood asthma and allergic disease. Transmitted and acquired risk factors collaborate to produce the phenotypic variation of a trait within the population. In this work, we have employed studies of twins to illustrate the relationship between twinship itself, early growth, and genetic and epigenetic factors with childhood asthma.

Twins have been suggested to be a high risk group for asthma. **Study I** was a large population-based register study of twinship in itself as a risk factor for childhood asthma. Asthma diagnoses and medication use among Swedish twins and singletons born 1993-2001 and 2005-2009 were compared before and after controlling for birth weight and gestational age. In the younger group, twins were at higher risk of developing asthma before controlling for perinatal factors – afterwards, twins were at lower risk of asthma in both age groups. This suggests that important mechanisms for asthma are shared between twins and singletons.

Low birth weight and rapid early growth have been shown to increase the risk of asthma. The aim of **Study II** was to describe the association between early growth and asthma in twins. Height and weight from 0 to 3 years of age were modelled in 2,874 twins. There was an association between later timing of maximum growth velocity and asthma both in terms of weight and height, although this relationship did not remain after controlling for birth weight or gestational age, which indicates that early postnatal growth may primarily be of interest as an extension of preceding foetal growth.

There is significant comorbidity between asthma and other allergic diseases. In **Study III**, we studied the influence of genetic factors on childhood asthma, hay fever, atopic eczema and food allergy. Using twin models and data from 25,306 twins, we concluded that asthma and all of the other allergic phenotypes were highly influenced by additive genetic effects. A pre-selected set of high-risk genetic variants were primarily associated with asthma or both asthma and hay fever (rs3771180 in IL1RL1).

Epigenetic factors have been suggested as a potential explanation for disease discordance within identical twin pairs. In **Study IV**, we analysed DNA methylation in whole blood of 708 twins with and without asthma using the Illumina 450k Beadchip. On the group level, 340 CpG sites were significantly associated with current asthma at 9-14 years of age, but these associations were not replicated within asthma-discordant pairs. Confounding by genetic factors or cell type composition in the samples seemed to be of importance, and should be considered as influences of potential importance in future epigenetic studies.

In conclusion, the work described in this thesis has combined the unique qualities of twin studies with data from population-based registers, interviews, and clinical examinations. The influences of transmitted (genetic) and acquired (twinship, early growth, and epigenetics) on childhood asthma were evaluated. From this selection, the transmitted factors proved to be of most importance to childhood asthma.

SVENSK SAMMANFATTNING

Hos barn såväl som ungdomar påverkas risken att utveckla astma av både arv och miljö. Eftersom tvillingar delar miljöfaktorer under uppväxten och en- och tvääggstvillingar har olika stora delar av sin genetiska kod gemensam, så ger tvillingstudier unika möjligheter att studera samspelet mellan ärftliga och förvärvade riskfaktorer.

Tvillingars astma i sig har dock varit ofullständigt studerad. **Studie I** syftade till att undersöka förekomsten av astma hos tvillingar jämfört med andra barn. Genom att länka data från Medicinska Födelseregistret, Patientregistret, och Läkemedelsregistret, kunde vi visa att tvillingar som grupp hade ökad risk att utveckla astma tidigt i livet, men inte efter sex års ålder. När resultaten korrigerades för att tvillingar ofta är födda tidigare och mindre än andra barn, så hade de istället *lägre* risk för astma. Att den förhöjda risken gick att förklara talar för att viktiga mekanismer för astma delas mellan tvillingar och andra barn.

Låg födelsevikt och snabb tidig tillväxt har tidigare setts öka risken för astma. I **studie II** var syftet att studera tvillingars tidiga tillväxt och sambandet med astma. Baserat på längd- och viktuppgifter från barnhälsovårdsjournaler beskrev vi 2874 tvillingars tillväxt från 0 – 3 år. De tvillingar som senare utvecklade astma visade sig ha sin maximala tillväxthastighet förlagd lite senare jämfört med de tvillingar som inte fick astma. När resultatet justerades för födelsevikt – och därmed den föregående tillväxten under graviditeten – fanns inte längre något sådant samband.

Astma och andra allergiska sjukdomar förekommer ofta tillsammans. I **studie III** ingick 25306 tvillingar, varav knappt hälften hade lämnat salivprov för genetisk analys. Vi beräknade heritabiliteten – den del av variationen av en egenskap som beror på gener – och konstaterade att den låg runt 70-80% för både astma och annan allergi vid 9-12 års ålder. Baserat på salivproverna kunde vi bekräfta att sex stycken genvarianter som tidigare kopplats till astma hängde samman med astma också hos svenska barn. Däremot var det endast en av dessa gener (varianten rs3771180 i IL1RL1-genen) som var kopplad till någon annan fenotyp (hösnuva).

I gränslandet mellan arv och miljö finns de epigenetiska faktorerna – som skulle kunna förklara hur enäggstvillingar, trots att de delar 100% av sina gener, ändå kan ha olika sjukdomar. I **studie IV** analyserade vi epigenetiska förändringar i blod hos 708 tvillingar med och utan astma. En del var diskordanta enäggstvillingar (den ena tvillingen i paret frisk, den andra sjuk). På gruppnivå hittade vi 340 olika epigenetiska varianter som var kopplade till astma – men det gick inte att se samma samband inom de diskordanta tvillingparen.

I de studier som ingår i denna avhandling har en unik sammansättning av datakällor och metoder förts samman för att studera ärftliga (genetik) och förvärvade (tvillingskap, tillväxt och epigenetik) riskfaktorer. Bland de faktorer som studerats här verkar de genetiska – som överförs mellan generationer – ha störst inflytande över astma i barndomen.

LIST OF SCIENTIFIC PAPERS

I. **Ullemar, V**, Lundholm, C, Almqvist, C. Twins' Risk of Childhood Asthma Mediated by Gestational Age and Birth Weight. *Clin Exp Allergy* 2015 Aug. 45(8);1328-36. doi: 10.1111/cea.12547

II. **Ullemar V**, Lundholm C, Magnusson PKE, Pershagen G, Lichtenstein P, Almqvist C. Delayed growth in twins with childhood asthma. *(Submitted)*

III. **Ullemar V**, Magnusson PKE, Lundholm C, Zettergren A, Melén E, Lichtenstein P, Almqvist C. Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy* 2016 Feb. 71(2):230-8. doi: 10.1111/all.12783

IV. **Ullemar V**, Karlsson R, Örtqvist AK, Lichtenstein P, Magnusson PKE, Almqvist C. Differential DNA methylation in childhood asthma: A case-control and discordant twin study. *(Manuscript)*

To date, two of the publications included in this thesis have been published in scientific journals (*Study I* and *Study III*). To receive e-mail alerts when the final versions of *Study II* and *Study IV* are published, you may sign up under the following link:

http://www.v.ullemar.se/subscribe/

RELATED PUBLICATIONS (NOT INCLUDED IN THESIS)

Ullemar V, Lundholm C, Örtqvist AK, Gumpert CH, Anckarsäter H, Lundström S, Almqvist C. (2015) Predictors of adolescents' consent to use health records for research and results from data collection in a Swedish twin cohort. Twin Res Hum Genet. 2015 Jun;18(3):256-65. doi: 10.1017/thg.2015.21. Epub 2015 Apr 22.

Almqvist C, Örtqvist AK, **Ullemar V**, Lundholm C, Lichtenstein P, Magnusson, PKE. (2015) Cohort Profile: Swedish Twin Study on Prediction and Prevention of Asthma (STOPPA). Twin Res Hum Genet. 2015 Jun;18(3):273-80. doi: 10.1017/thg.2015.17. Epub 2015 Apr 22.

Protudjer JLP, Binnmyr J, Grundström J, Manson M, Marquardt N, Säfholm J, **Ullemar V**. (2015) Allergy trainees' perspectives on career opportunities: Results from a trainee-organised retreat. Allergy. 2015 Nov;70(11):1353-5. doi: 10.1111/all.12690. doi: 10.1111/all.12690.

Almqvist C, Olsson H, **Ullemar V**, D'Onofrio B, Frans E, Lundholm C. (2015) Association between parental age and asthma in a population-based register study. J Allergy Clin Immunol 2015 Oct;136(4):1103-5.e2 doi: 10.1016/j.jaci.2015.04.006.

Movin M, Garden FL, Protudjer JLP, **Ullemar V**, Svensdotter F, Andersson D, Kruse A, Cowell CT, Toelle BG, Marks GB, Almqvist C. The impact of childhood asthma on growth trajectories in early adolescence: Findings from the Childhood Asthma Prevention Study (CAPS). Respirology *(2016, In press)*

CONTENTS

LIST OF ABBREVIATIONS

ADRB2	Beta-2-adrenergic receptor gene
A-TAC	Autism-tics, AD/HD and other comorbidities inventory
BMI	Body mass index, kg/m^2
BW	Birth weight
CATSS	Child and Adolescent Twin Study in Sweden
CI	Confidence interval
CNV	Copy-number variation
CS	Caesarean section
DAG	Directed acyclic graph
DNA	Deoxyribonucleic acid
DOGSS	Developmental Outcomes in a Genetic twin Study in Sweden
DOHAD	Developmental origins of health and disease
DZ	Dizygotic, fraternal (non-identical) twin pairs
EDTA	Ethylenediaminetetraacetic acid
ER	Eosinophil to leukocyte ratio
EWAS	Epigenome-wide association study/studies
FDR	False discovery rate
GSDM	Gasdermin gene(s)
GWAS	Genome-wide association study/studies
HR	Hazard ratio
ICD	International Classification of Disease
ICS	Inhaled corticosteroids
IgE	Immunoglobulin E
IUGR	Intrauterine growth restriction
Limma	Linear Models for Microarray Data
LRT	Likelihood ratio test
LTRA	Leukotriene receptor antagonist
MAP3K6	Mitogen-activated protein kinase 6
MAP3K7	Mitogen-activated protein kinase 7
MBR	Medical Birth Register
MZ	Monozygotic, identical twin pairs
NBHW	National Board of Health and Welfare (Swedish: Socialstyrelsen)
NPR	National Patient Register

NR	Neutrophil to leukocyte ratio
OR	Odds ratio
ORMDL3	ORMDL sphingolipid biosynthesis regulator 3
PIN	Personal identification number
RR	Relative risk
SGA	Small for gestational age
SITAR	SuperImposition by Translation and Rotation
SNP	Single nucleotide polymorphism
SPDR	Swedish Prescribed Drug Register
STOPPA	Swedish Twin study on Prediction and Prevention of Asthma
STR	Swedish Twin Registry
WBC	White blood cell
WHO	World Health Organization

1. FOREWORD

While the purpose of a thesis summary is fairly clear – to provide a context for and discuss the scientific work carried out by an individual over the course of a PhD education program – identifying the target audience is a more delicate matter. Many doctoral theses are only truly read by a handful of people (the PhD candidates themselves excluded – as the author of this work, I doubt I will ever acquire a truly unbiased opinion of it) – which one might argue is a very anticlimactic outcome of all the effort it took to achieve it. I can, of course, not make anyone read this; nor would I truly wish for my words to be forced upon anyone. But I can aim to make it an enriching experience to try.

On that note, I have tried to write a thesis in the spirit of open communication, honesty, and filled with the true enthusiasm I have felt while carrying out the work presented here. I hope it shows. And in that spirit – because science truly is fun – I hope that once you've read it, however small a part, if there are any questions or thoughts born in your mind as you do so that you will bring them to me. I'll be there.

Now: let's talk science.

2. INTRODUCTION

2.1 WHY STUDY ASTHMA?

We live, in part, because we breathe. The art of respiration has been developed by evolution over millions of years.[1] The human lungs, with their vast network of airways and alveoli, are ultimately a consequence of evolution and development; a biological machinery with the vital purpose of providing our bodies with the oxygen we infallibly need.[2] Perhaps it is for this reason that respiratory diseases can be so utterly unsettling for those affected.

Asthma affects hundreds of millions of people across the globe. The latest numbers from the World Health Organization, WHO, reveal an estimated total of 235 million people with asthma worldwide.[3] A significant proportion of them are children; asthma is the most common non-communicable disease in this age group. As such, it is an important factor in worldwide childhood morbidity.[4,5]

Yet despite the continued high prevalence of this disease[6] and a substantial number of identified factors which are associated with an increased risk of asthma within a population,[7] piecing together these many different clues into a whole is an on-going challenge.

Though one may find several pieces of a puzzle, identifying their proper place in the design is a challenge in itself. What this project aims to accomplish is not to finish the puzzle; I trust that it will continue to occupy my colleagues and myself for decades to come. The overarching aim for this thesis is to highlight a few particular features, in an attempt to shed new light on these well selected conundrums within the field.

Before we get there, I will take you through a few concepts concerning epidemiological research in general, and genetic and early life epidemiology in particular.

2.2 A NOTE ON DEFINITIONS

The title of this thesis contains two keywords which merit clarification: **transmitted** and **acquired**. These are applied to concern specific elements hypothesised to be risk factors for childhood asthma.

Within the scope of this thesis, **transmitted** risk factors are mainly the inheritable: specific genetic variants or heritability overall, as in *Study III*. Thus, the process of transmission considered here is primarily that which occurs between generations.

Acquired risk factors refer to things a person may become exposed to or experience as a consequence of an environmental influence. In this work, the acquired risk factors discussed are infant growth (*Study II*) and epigenetic modifications (*Study IV*). It is also possible to view the epigenetic mechanisms as a bridge between the concepts of transmission and acquisition, in the sense that they act on the genetic factors but mediate environmental effects.[8]

Whether **twinship** (the exposure of interest in *Study I*) belongs in the transmitted or acquired category is less obvious. One might argue that as monozygotic (identical) twinning occurs after conception, it is acquired. On the other hand, dizygotic (fraternal) twinning, which happens when two zygotes are conceived at the same time, more commonly occurs in some families[9] and could thus be considered a transmitted trait. While the experience of being a twin comes with particular features of intrauterine environment and the presence of a sibling of equal age, the acquired features of twinship may be of more importance.

But the beauty of the thesis title does not lie in disentangling exactly which risk factor belongs where – rather in illustrating the position of the subject: balancing between the roles of genes and environment.

2.3 EPIDEMIOLOGY AS A FIELD OF RESEARCH

Epidemiology is the discipline of medical science dedicated to the study of how a set of factors (often called exposures) are associated with a certain set of other factors (the outcomes). If people who have experienced the exposure on average more often get the outcome, we say that people with the exposure have an increased risk of developing this outcome. It is important to remember, however, that this is not an absolute.

Indeed, on an individual level, interpreting a statement of risk tells you very little. The mere presence of a well known risk factor does not guarantee that a person will go on to develop the outcome – only that they are statistically more likely to do so compared to someone who is unexposed. As with everything else in life, the findings from new epidemiological studies should be viewed in balance with the already available information. For application in clinical settings, findings from observational studies hold a lower level of evidence compared to randomised clinical trials,[10] but they are often an excellent starting point for answering questions arising in the clinic. Further, they may provide clues towards prevention of future cases of disease.

However, for this purpose, the strength of the estimated association matters. Even a strong association in itself means little for the burden of disease in the population at large if the prevalence is low.

Further, the fact that there is an association between an exposure and the outcome does not guarantee that the association was caused by the direct effect of the exposure on the outcome. Epidemiologists are ever concerned with avoiding the side effects of **confounding** – when a third factor associated both with the exposure and the outcome obscures the true relationship between the two, either making it seem as if there is one while there truly isn't, or by hiding a true association.[11] Additionally, if a study is cross-sectional – i.e. the exposure and outcome of interest have been measured at the same time – **reverse causality,** the possibility that it is the outcome that has caused the exposure and not the other way around, can sometimes not be excluded.

Finally, the results drawn from a study are only as good as the accuracy of measurement of the exposure and outcome will allow. Sometimes, observational studies are studies of the most suitable proxies of the true phenomenon. **Measurement errors** may arise when there is some disagreement – whether random or systematic – between what is measured and the true value.[12] Additionally, **information bias** may result from misreporting of exposures or outcomes (such as by a disturbance in recollection of the exposure dependent on outcome status – referred to as recall bias[13]) and **selection bias** from performing a study within a group into which the exposure and outcome were factors in representation.[14,15] All of these – and more – potential sources of errors and biases should be evaluated in the interpretation of results from epidemiological studies.

Ultimately, while it is important to be aware of the general limitations of observational studies, they also have significant strengths. Some things cannot be randomised, and epidemiological studies – carried out within population-based registers or large well-characterised cohorts – can provide uniquely powerful settings in which to apply intelligent designs to investigate the reality of the world that surrounds us.

3. BACKGROUND

3.1 ASTHMA AND ALLERGIC DISEASES

3.1.1 What is asthma? Phenotypes, pathophysiology, treatment, and immunological aspects

3.1.1.1 Asthma phenotypes

While asthma is common during childhood, the characteristics of disease vary within this time span as well as between childhood and adulthood.[16] Attempts have been made to describe these different expressions of the disease in terms of asthma and/or wheezing **phenotypes**. In 1995, one of the most widely acknowledged early attempts was made by Martinez.[17] Since then, several attempts have been made to continue to build on either this or similar classification strategies.[18-24]

Briefly, three different aspects are commonly considered in some way in most of these attempts: the **age** at which the disease is experienced, the **trigger factors**,[25,26] and the disease **severity**.[27] The age at which a child experiences asthma can be described in several different ways. One important dimension is the age of onset; conversely, it is also of interest until which age symptoms last.[23] It is worth noting, however, that while there is a general agreement that some sub-classification of asthma is likely warranted, the possibility of making these distinctions within a given study depends on the level of detail in the collected information. From a clinical perspective, identifying factors that may lead to improved diagnostic methods or personalised treatment is probably of greatest interest.[28]

A recurring argument within the field is whether the children who start wheezing at a very early age – many of whom do so in association with respiratory infections such as respiratory syncytial virus (RSV) or adenovirus infections – but cease to do so within a few years should be considered asthmatic. Today, many of them are referred to as '**early transient wheezers**', and there is often discussion regarding whether to include these individuals within the asthma group in epidemiological studies, or to exclude them so as to retain the specificity of the outcome definition. Symptomatologically, all conditions that feature wheezing and obstruction among children lead to childhood respiratory morbidity, and may share common causes. Still, there has been some evidence that genetic variation differs between some early wheeze phenotypes,[29] which would speak in favour of studying them separately to the largest extent possible. On the other hand, early wheezers who continue to wheeze until at least 18 months of age may continue to show signs of airway inflammation – and, possibly, asthma – well into adolescence.[23]

Due to the large overlap of respiratory symptoms, regardless of their persistence, it makes sense to still consider asthma as primarily one disease – but a heterogenous one. When translating results to clinical settings, the limitations of the phenotype definition under study should be acknowledged.

3.1.1.2 Pathophysiology and treatment

Physiologically, asthma is characterised by **hyper-reactivity** in and **inflammation** of the airways. Hyper-reactivity leads to constriction of the airways in response to an irritating agent (such as allergens, cold air, and exposure to tobacco smoke) owing to contraction of smooth muscle cells within the bronchial wall. Additionally, secondary to the inflammation, there is often excess production of mucus within the airways of people with asthma.[30] Alone or in synergy, these factors can result in **bronchial obstruction** and be very troublesome.[31]

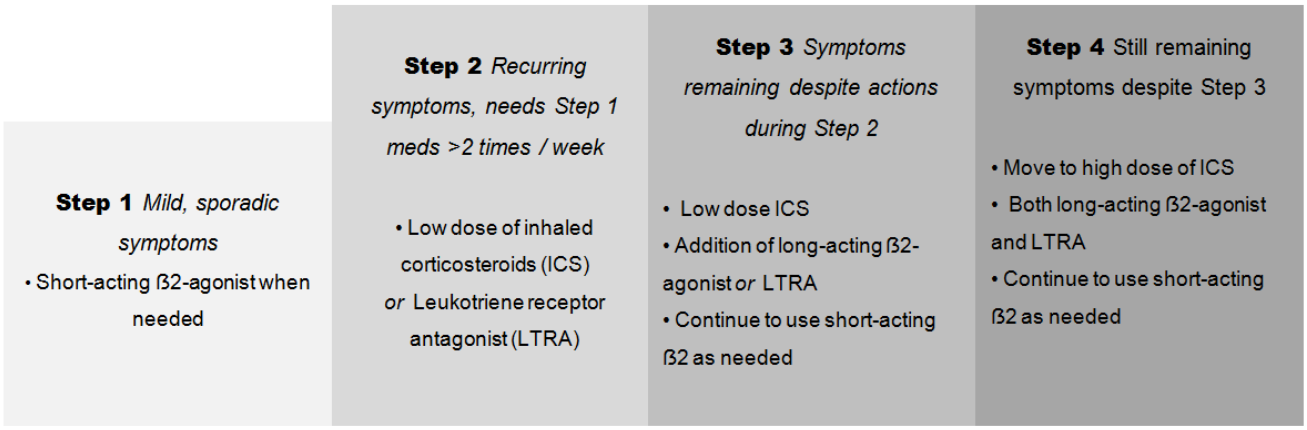

Figure 1 *– Swedish treatment guidelines for continuous treatment of childhood asthma after 6 years of age. Adapted from the Swedish Paediatric Association, section for allergology (www.barnallergisektionen.se)*

Treatment of childhood asthma targets both of these dimensions. An overview of asthma symptoms and the Swedish Paediatric Association, section for allergology treatment ladder for different types of medication is presented in *Figure 1*, above. Swedish practice is in line with international recommendations.[32]

Briefly, for sporadic or mild symptoms (Step 1), short-acting beta-2-agonists are prescribed and used primarily when the patient anticipates or has already developed symptoms. The purpose of this medication is to relax the overly constricted airways by selectively activating the beta-2-receptor (which mediates smooth muscle relaxation). Long-acting beta-2-agonists act according to the same principles, but are introduced at later treatment steps and always together with ICS (Step 3 and onwards). Long-acting beta-2-agonists are taken regularly and have longer half-life than short-acting beta-2-agonists.

The chronic inflammation of the airways associated with asthma is treated with inhalations of corticosteroids (ICS). This is a standing prescription for many individuals with asthma, (Step 2 and up). Many children only need to use ICS intermittently, such as when exposed to pollen and pets, during respiratory tract infections or conjunction with to exercise. Additionally, leukotriene receptor antagonists (LRTA) may be used as an alternative to ICS (Step 2), or as an addition to the aforementioned treatment regimen (Step 3 and 4).

3.1.1.3 Immunological aspects

Asthma is both associated with and mediated by cells of the immune system. A key discussion has long been about the specificity of the T-cell response in the disease. Briefly, the adaptive immune system includes several cell types that mediate the acquired immunity of the host. The main cell family making up the adaptive immune system are the lymphocytes (a subset of leukocytes), which in turn can be divided into T, B and natural killer (NK) cells.[33] B cells produce antibodies, whereas T cells are the designated actors who clear out pathogens. NK cells are primarily associated with the innate immune response.

Traditionally, T-helper cell responses have been seen as weighted in several disease patterns.[34] For instance, in auto-immune diseases, where the immune system starts reacting against the body's own cells, the response has long been characterised as Th-1-weighted. In contrast, in asthma and allergic diseases, where the pattern is instead characterised by an overreaction to external stimuli such as allergens or other irritants, the response has been seen as Th-2-dominant. Recently however these responses have proved to require more complex descriptions,[35,36] and genetic variation along the T-helper cell pathway may be involved in disease susceptibility.[37]

3.1.2 Comorbidity of asthma with hay fever, eczema, and food allergy

Asthma displays significant comorbidity with several diseases within the allergic and atopic spectrum.[38-41] The phenomenon called the "atopic march" gets its name from the idea that there is a natural progression of allergic diseases (primarily atopic eczema, asthma and hay fever) during childhood.[42-46] In practice a child may develop these phenotypes in different orders,[47] but comorbidity[38] and IgE levels[48] increase with age. Underlying this significant comorbidity, the predisposing risk factors may be shared (such as the large proportion of genetic components that are common to both asthma and hay fever[49]) or directly opposing. One example of the latter was shown in a pair of studies within the Swedish Twin Registry, in which low birth weight increased the risk of asthma,[50] but high birth weight increased the risk of atopic eczema.[51] Additionally, exposure to environmental tobacco smoke in the postnatal environment is associated with asthma, but less strongly associated with childhood eczema.[52] Maternal smoking during pregnancy does not appear to be associated with eczema.[53] This tendency for risk factors for asthma and allergic diseases to overlap, but also contain differences with respect to asthma vs. eczema or hay fever, continues in adulthood.[54] Comorbidity may also affect asthma symptoms. For example, children with some food allergies who also have asthma are at increased risk of hospitalisation and require more steroid treatment for their asthma.[55]

Thus, comorbidity between asthma and allergic disease is both significant and of potential clinical consequence. The specific nature of shared risk factors for these diseases merits further investigation.

3.2 EARLY LIFE EPIDEMIOLOGY

3.2.1 What happens in utero doesn't stay in utero

The hypothesis of developmental origins of health and diseases (DOHAD) was first discussed in the work of David Barker in the context of the effects of restricted foetal growth on adult cardiovascular health.[56,57] Later, the hypothesis has been extended to the study of several other phenotypes, including respiratory diseases.[58] In the case of asthma, the exact biological mechanisms by which conditions during foetal life may affect later respiratory health are not clearly established. There are however several established exposures originating during gestation that have effects on later health. These include restricted foetal growth,[50,59-64] premature birth,[61,65,66] and intrauterine exposure to metabolites associated with maternal smoking during pregnancy.[67]

3.2.2 Early life exposures and asthma

Postnatal early life environment contains several different factors which have been hypothesised to affect asthma in childhood. Many of these supposed relationships have amounted conflicting evidence in their favour and have been the subject of much discussion within the scientific community. Some of the more stable associations have been a protective effect of belonging to a group of siblings[68,69] and early exposure to animals,[70] as well as increased risks of asthma associated with exposure to environmental tobacco smoke early in life[71] and certain viral infections[72] – although in the latter case reverse causality has been difficult to exclude. The same has been the case for early exposure to antibiotics, although much of the initially shown association between antibiotic exposure and asthma has recently been shown to have been due to familial confounding.[73] Further, early growth and effects mediated via genetic and epigenetic mechanisms have been suggested to be of importance. These are described in the following sections.

3.3 GROWTH AND ASTHMA

Human growth is often described as divided into stages during which there are defining characteristics and features. But the borders between these stages are not as set in stone as they may seem – in reality, there is a continuum of changes and features may overlap. Still, the distinctions serve a purpose in illustrating some particularly important characteristics of each phase.

3.3.1 Phases of early life growth

Some of the most essential growth of our lives occurs in utero, where we transform from a single fertilised oocyte into a small human being of diverse cellular composition – within a matter of months.

Intrauterine growth restriction (IUGR) may be caused by any number of factors, but usually results in a child being small for gestational age (SGA).[74] This is principally different from a child who is born small because they were not carried to term. In other words, two infants of

equally low birth weight may be small or appropriate for gestational age depending on the length of gestation – thus birth weight alone is not a sufficient proxy for SGA. In twin pregnancies, IUGR is more common than for singletons – but may affect only one of the twins.[75] Previous studies of low birth weight and asthma have generally shown that low birth weight is associated with an increased risk of asthma in childhood.[50,59-64]

Although exceedingly rapid **growth in infancy** may be considered an adverse growth pattern, the natural state of growth during the first two years of life is rapid in itself. Indeed, during the first year, a child is said to on average triple their weight.[76] Still, children who are born with low birth weight or small for gestational age may experience a phenomenon referred to as **catch-up**.[77] This type of growth pattern has been linked to obesity later in childhood, and some theorise that it may be associated with activation of inflammatory pathways.[78] On the other hand, early rapid weight gain also occurs in infants who have not been born with low birth weight – although low birth weight is a strongly predisposing factor, feeding patterns and child's gender are also of importance.[79] Because it is possible that early infant growth occurring in response to an earlier restriction (catch-up) may be biologically different from differences occurring in response to current environment (such as would be the case with responses to feeding patterns or other modifiable factors), it may be valuable to distinguish between these two principally different patterns when it comes to their relation to childhood health outcomes.

3.3.2 Previous studies of childhood growth and asthma as the outcome

Initially, many studies of childhood growth and later or concurrent asthma did not study the growth process per se: rather, body mass index (BMI, kg/m^2) was a commonly used exposure. Although this is a natural endpoint of childhood growth processes and may thus be considered a proxy of growth, several different growth patterns may result in the same value of BMI.

Nevertheless, in terms of asthma, it was generally found that high **childhood BMI** held a positive association with disease.[80-88] A major weakness of many of the BMI studies was that they were largely cross-sectional in nature – meaning it was not possible to exclude reverse causality.

Soon, studies of **BMI trajectories**, or transitions between categories of BMI between different ages, became a new addition.[89-94] These had the advantage of illustrating a direction of childhood growth between two or more time points, and were able to show that moving upwards in terms of BMI during childhood – or having a persistent high BMI – was associated with asthma. Early high BMI that later fell to normal levels did not seem to be associated with an increased risk.[89,90] Similar patterns were later replicated in studies using latent growth mixture models[95] and BMI gain Z scores.[96] It may be important to keep in mind that changes in BMI during childhood may not be associated with asthma later in life.[97]

Still, studies of BMI on its own could not capture nuances in growth that primarily occur in terms of **changes in weight or height separately**. Using longitudinally collected measures

from early childhood, other studies began to show that rapid increases in weight were associated with asthma or wheeze,[98-110] and one study found that slow weight gain before 2 years of age was protective against asthma symptoms.[111] Concerning height gain, null associations[99,101,102,104,105,108,110] were most commonly reported, although one study found a negative association between rapid height gain between 3-7 years and asthma at 8 years of age.[109] Some studies applied more **complex statistical models** to describe specific features of growth patterns, rather than making calculations based on absolute increases or rates.[102,104,107,110,112] Generally, significant differences concerned growth that occurred during the first year of life.

But as outcomes of childhood growth[113-115] and asthma[116] are both highly dependent on genetic effects, it is possible that some of the aforementioned associations could be partially confounded by genetic factors.

3.4 GENETIC AND EPIGENETIC EPIDEMIOLOGY

3.4.1 The Human Genome: genetic and epigenetic variation

A key principle to understanding deoxyribonucleic acid (**DNA**) structure is base pair combinations. The DNA backbone, the spine of the molecule, exists in 2 copies. Either of these can be used to re-create the other. This is due to the fact that DNA consists of a sequence of molecules: the nucleotides. These are adenosine (A), thymine (T), cytosine (C) and guanine (G).[117] Due to the chemical structure of these bases, they can only be paired together in two ways within the DNA helix; A with T and C with G.

The vast majority of the DNA sequence is identical within the human population. Yet in the half percent of the human genome that varies between individuals lies the key to our individuality.[118] There are several different kinds of genetic variation, one of the most readily studied being the **SNP: single nucleotide polymorphism.**[119] An SNP is a position within or near a gene in which different people have different alleles or genetic variants. At this position, some people may have one nucleotide, say A, but others another, say C. Due to the inheritance of one chromosome from each biological ancestor, a person may have different alleles of the same SNP on each chromosome.

Epigenetic modifications are reversible molecular changes to the DNA molecule, which do not alter the genetic code; they do however affect its application.[120] The most common type of epigenetic modification is **DNA methylation,** *Figure 2*. DNA methylation is performed by enzymes of the DNA methyl transferase family, which act to attach a methyl group (CH3-) to cytosine (C).[121]

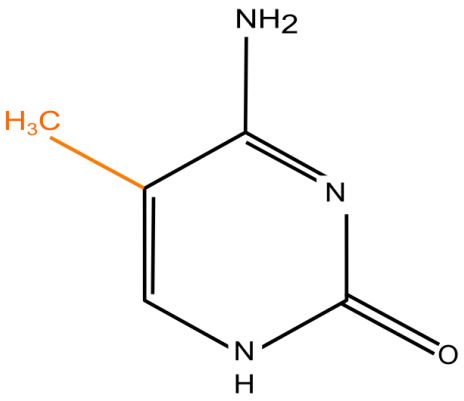

Figure 2 *Chemical structure of a methylated cytosine molecule (5-Methylcytosine). The methyl group (-CH_3) (orange) has been exchanged with a hydrogen molecule by the DNA methyltransferase enzyme.*

DNA methylation most often occurs in connection to a so-called **CpG site**, which is any specific location along the DNA molecule where a cytosine base (C) is immediately followed by guanine (G).[122] Because DNA methylation occurs specifically at the CpG site, there needs to be a CpG site for it to occur at. Thus, individual genetic variation has at least one obvious association with whether a particular locus is methylated or not. This would be expected to have a certain impact on between-individual variation in DNA methylation,[123,124] but studies have also shown that there are more profound ways in which the gene sequence may influence the methylation patterns, which go beyond this.[125,126] All of these mechanisms could be the underlying causes of so-called **genetic confounding**: the risk run by any study of epigenetic mechanisms which is carried out in a population of genetically unrelated individuals. Additionally, DNA methylation varies by cell type, and the tissue choice should be taken into account when interpreting results of DNA methylation analyses.[127]

3.4.2 GWAS and EWAS

Some of the first studies aiming to study **genetic variation** were so-called candidate gene studies.[128,129] These studies specifically targeted genes which were either known to be involved in a particular pathway relevant to the disease, or otherwise hypothesised to be of importance for the phenotype in question, leading to disparate findings – many of which later proved difficult to replicate.[130,131] One possible explanation for this is that the process of selecting an appropriate candidate gene propagates a bias based on the previous knowledge used to select it.

Later, pre-designed assays testing a wide variety of loci at the same time were used. At first these chips were relatively small, testing a few tens of thousands of variants; later, they would test hundreds of thousands to millions of loci at the same time. Such studies are referred to as genome-wide association studies (**GWAS**).[132,133] Although these approaches are often referred to as 'hypothesis-free', this is only true to the extent that scientists are looking for anything within this pre-defined subset of variability – a fishing expedition, to be sure, but in a pre-selected lake. As a result, different research groups have become accustomed to pooling their data (creating a sea out of the sum of several lakes, if you will) in large consortia. The results of these pooled analyses are often considered to be more reliable, as the power to detect even small associations becomes greater.[134] A commonly used argument against these studies is the fact that heterogeneity in study designs may lead to a dilution or bias of the true results.

The technology for detecting epigenetic variation has developed along many of the same pathways, but a few years later and at slightly more rapid pace (after all, it's easier to copy an evolution of concepts than it is to invent it the first time around). Currently, we are in the era of **EWAS** – epigenome-wide association-studies and their subsequent consortia. As for GWAS, EWAS are now possible to perform using previously created arrays that test thousands of pre-defined loci at the same time. The Illumina 27k,[135] 450k[136,137] and only recently 850k[138] DNA methylation chips test methylation at tens to hundreds of thousands of CpG loci. This technology however comes with some limitations – of particular note is the

risk of **batch effects.**[139] These occur when systematic differences in signal strength occur between sample plates or chips as an artefact of the method. To avoid a resulting bias, sample allocation within and between plates must be performed with care.

3.4.3 Asthma and the genes

Asthma is a highly heritable disease [49,116,140-145] but the specific features of this heredity – i.e. which genetic variants are primarily responsible – has been established mainly during the recent decade.[146-152] There have been several large GWAS efforts in which asthma has been the phenotype of interest.[146-148,153-165] Some of these have included both adults and children,[147,148,157,161,162,166] or only one of these age groups.[153-155,158-160,163,164] Due to the differences in asthma phenotype by age, there is a possibility that different genetic variants are of importance for asthma at different ages.[167]

One of the most significant meta-analyses of asthma GWAS to date was carried out by the GABRIEL consortium and included data from over 20 materials, featuring both adult- and childhood-oriented cohorts.[146] The results confirmed some but did not replicate all findings from previous studies (including both smaller GWAS and candidate gene studies). While it is possible that some of the failed replications were due to false positives in the original studies, the diversity of the definition of asthma diagnosis itself (one of the most common criticisms towards the GABRIEL consortium) could also be part of the explanation.[154]

The numerous discovery efforts and their consecutive replication of some of each other's results have resulted in clear indications regarding the importance of especially regions on Chromosome 2, 5, 6 and 17.[146,147,155,160,161,164] For childhood asthma, the ORMDL3 (or gasdermin A / gasdermin B) locus on chromosome 17 featured particularly strong associations.[146,168] Still, studies differ in terms of included age ranges, and different genetic variants may be of importance in other populations. It remains worthwhile to pursue targeted replication efforts within specific populations and age ranges, particularly in childhood.

3.4.4 Asthma and the epigenome

Epigenetic studies of asthma have also moved into the era of epigenome-wide array studies,[169-174] but so far these present an inconclusive image. Results from collaborative consortia studies are – as of now – yet to be reported, and available studies have not disclosed the full results of their analyses. Within the top findings of statistical significance from these studies (including roughly 100 specific CpG loci),[169-174] very few findings have been replicated. Specifically, these concern findings within the B2-adrenergic receptor gene (ADRB2)[169,175] and the mitogen-activated protein kinases 6 and 7 (MAP3K6[171] and MAP3K7[169]). Methylation in the ORMDL3 region may be associated with asthma control,[176] but the finding has not been replicated within discovery studies of asthma. In general, the overlap between findings in terms of genetic and epigenetic variance is limited. This speaks in favour of continuing the epigenome-wide approach, and also hints that the interplay between genes and environment may feature interactive effects between genomic regions.

As was the case with genetic variation, it's possible that childhood asthma carries particular features also in terms of epigenetic variation. Additionally, genetic or cell type confounding has not previously been accounted for in most epigenetic studies of asthma: there are only two previous studies using twin designs.[172,177]

3.5 WHY TWINS?

3.5.1 Shared genes and environment – assumptions, advantages, and reasons for investigation

For the genetic scientist, twins hold a special place of importance. The reason begins with the core of genetics: the DNA sequence, and the distinction between **monozygotic** (identical or MZ) twins and **dizygotic** (fraternal or DZ) twins. It is generally assumed that monozygotic twins share 100% of their DNA code and that DZ twins, like any other pair of full siblings share, on average, 50% of their **segregating genes**. As a result, MZ twins are always of the same biological sex, whereas DZ twins may be same-sexed (DZ-ss) or opposite-sexed (DZ-os). It should be noted that this assumption is not necessarily 100% correct, as post-division point mutations[178] or copy number variations (CNVs) within the genome may occur also within MZ twin pairs.[179] However, recent studies indicate that such differences are rare.[180,181]

Beyond the genetic, all twin pairs share their environment: both that in utero and, usually, the one in which they are raised. It is assumed that this sharing occurs to an equal extent across zygosities (also referred to as the **equal environments assumption**). But twin pregnancies are particular in the sense that they require two children to develop simultaneously in the womb. In practice, this may be associated with sharing both of the placenta (monochorionicity) and the amniotic sac (monoamniosity), depending on during which day of gestation twinning has occurred. Generally, these features are more common in pregnancies with MZ twins, but monochorionic DZ twins exist.[182] Chorionicity may have a small effect on both birth weight[183] and birth weight discordance within twin pairs.[184] While the magnitude of the long-term impact of these intrauterine differences is largely unknown, their importance should be seen in relation to several other important features that twins are known to share, such as their age, siblings, parents, and usually home environment.

Other specific features of twin pregnancies – such as on average shorter gestations and lower birth weights compared to singletons[185] – holds potential importance for twin studies. When these factors are also associated with a phenotype under study, it may be of interest to investigate whether twins seem to differ from singletons in terms of this phenotype.[186]

Together, the assumptions of shared genes and environment within twin pairs are the core of three methods applied within this thesis: heritability estimation (*Study III*), co-twin control analysis (*Study II*), and within-pair comparisons of DNA methylation (*Study IV*).

3.5.2 Heritability

Heritability is a term for the **proportion of the total variation** of a disease or phenotype that is due to genetic factors. In twin models, heritability, also known as **additive genetics**, is usually symbolised by the factor A. Together with **non-shared environment** (E) and **shared environment** (C) or **genetic dominance** (D), it adds up to the sum of the total variation.[187] *Figure 3* presents a simplified structural representation of such a model.

Heritability estimates may differ between populations based on several different factors.[188] One explanation is if genetic factors are truly of different importance in these populations – but it could also be that the relative importance of environmental factors is different. One example might be if maternal smoking during pregnancy – which has been associated both with childhood asthma[189] and early wheeze[190] – is a bigger concern in one population versus another.

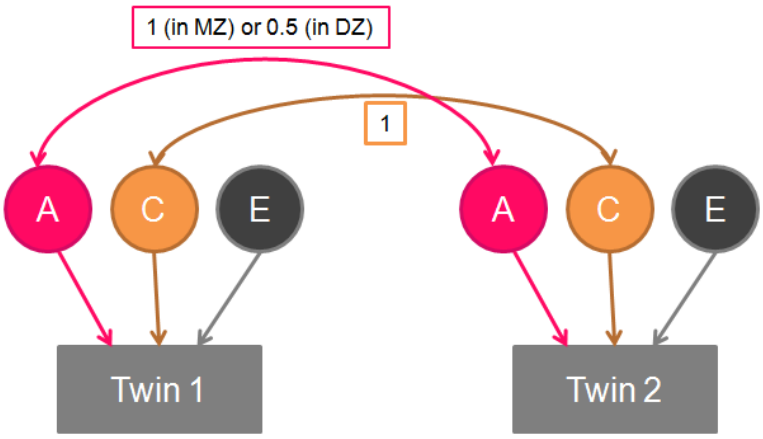

Figure 3 *Structural representation of the ACE twin model, displaying the assumed correlations (boxes adjacent to the lines) between factors additive genetics (A), shared environment (C), and non-shared or unique environment (E) within a twin pair (Twin 1 and Twin 2).*

In addition, heritability estimates may vary over time, between age groups,[191] or for different phenotypes of the same disease. Adding to and updating available estimates of heritability can be a worthwhile effort that gives an indication of the current impact of genetic versus environmental factors for a phenotype and target group of interest.

3.5.3 Within-pair and co-twin control analyses

Whereas twin models are used to quantify the size of the total impact of genetic versus environmental factors on a phenotype, within-pair and co-twin control analyses are applied to a specific association between exposure and outcome.[187] A prerequisite for both of these methods is that the included twin pairs are discordant (differ in their status of the trait within the pair) for both exposure and outcome – i.e. constitute **informative pairs.** Because this can be rare, within-pair analyses often include a relatively small number of individuals even when they originate from large twin cohorts. That said, these methods can present unique opportunities of discerning whether genetic or environmental confounding is of importance for the relationship under study – or highlight findings controlled for these factors. A further discussion of these methods is available in the *Methods chapter, section 5.3.9.*

4. AIMS

The intent of this work is to study if and how a collection of transmitted and acquired factors influence childhood asthma. The thesis comprises four studies – their separate specific aims are as follows:

To . . .

1. Investigate the association between twinship and asthma, as well as to determine whether gestational age and birth weight mediates the potential association. (*Study I*)

2. Test whether parameters of early growth were of importance for childhood asthma, and if genetic or familial confounding partially explains any such associations. (*Study II*)

3. Present new heritability estimates of asthma, hay fever, eczema and food allergy in mid-childhood. Secondly, to investigate whether genetic variants previously linked to childhood asthma are associated with this or related phenotypes in Swedish twins. (*Study III*)

4. Compare DNA methylation in whole blood between twins with and without asthma, in the full STOPPA cohort as well as within asthma-discordant MZ and DZ twin pairs. Parallel analyses concern confounding by cell type. (*Study IV*)

5. METHODS

5.1 DATA SOURCES

Regularly recording and following up on statistics of the health of the Swedish population is one of the main activities of the National Board of Health and Welfare (Swedish: Socialstyrelsen).[192] This serves both administrative and informative needs. For example, what later became the Medical Birth Register was originally started in response to a rise in congenital malformations that followed after pregnant women were prescribed the drug Neurosedyn (Thalidomide) during pregnancy.[193] In addition to such **population-based registers**, cohort studies constitute another valuable source of information regarding public health. These **cohort studies** often originate from efforts of individual research groups or collaborations, and include targeted collection of data specifically for research purposes.

Using the unique Swedish personal identification number (PIN), data from different sources may be linked together.[194] This linkage allows for studies of intricate exposure–outcome relationships, with the potential to account for familial aggregation, socioeconomic status, and perinatal factors. It is important to note that although the resulting data can be extensive, it is only analysed in de-identified format. Individual researchers cannot connect the coded personal ID numbers created to the original PIN or identities of the study participants.

This chapter introduces the population-based registers and cohorts featured within this thesis.

5.1.1.1 The Swedish Twin Registry (STR) and relevant cohorts

The Swedish Twin Registry (STR) is a collection of several age-based cohort studies.[195] The relationships between the STR cohorts used in the work described in this thesis are shown in *Figure 4.* Further details regarding the cohorts are available below.

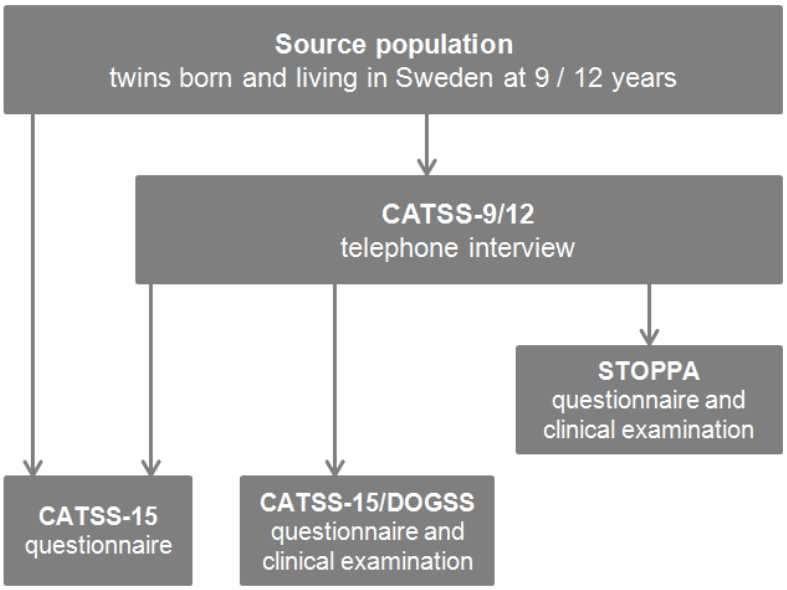

Figure 4 *Relationship between the STR cohorts included in this thesis: CATSS-9/12, CATSS-15, CATSS-15/DOGSS, and STOPPA.*

5.1.1.2 The Child and Adolescent Twin Study in Sweden: CATSS-9/12

Since 2004, twins born and living in Sweden are invited to the Child and Adolescent Twin Study in Sweden (**CATSS**) following their ninth birthday.[196] During the first three years of data collection, the STR also invited twelve-year-old twins to CATSS.

Twins born 1996 and onwards have always been nine years of age at inclusion in CATSS. The oldest twins in CATSS were born in 1992 – the youngest in 2007.

Participation in CATSS starts with a telephone interview with the twins' parents. The extensive interview includes questions on child's health, medication use, living situation, perinatal factors, and current weight and height, among other traits. A module including questions regarding the twin pair's physical similarities is the basis for an algorithm-based assessment of zygosity. Later, twins have also been offered to determine their zygosity by DNA testing of saliva samples.[197] To date, approximately n=28,800 twins have participated in CATSS-9/12.

Study II, *Study III* and *Study IV* are all in some way connected to the CATSS-9/12 cohort, making it one of the most central data sources in this thesis.

5.1.1.3 CATSS-15 , CATSS-15/DOGSS, and child growth records

CATSS-15 and CATSS-15/DOGSS (Developmental Outcomes in a Genetic twin Study in Sweden) were both follow-up studies of CATSS-9/12 at 15 years of age.

The main difference between the two studies concerns the invitation of study participants. The only requirement to be invited to CATSS-15 was to have qualified for an invitation to CATSS-9/12. To actually have participated was not a prerequisite. In contrast, twins invited to CATSS-15/DOGSS were selected based on screening for neurodevelopmental disorders in CATSS-9/12 using the Autism-tics, AD/HD and other comorbidities inventory (A-TAC) inventory.[198] The full CATSS-15/DOGSS cohort also included co-twins of the screened twins (even if these had not screened positively themselves), as well as controls.

Although CATSS-15/DOGSS also included a clinical examination with an in-depth evaluation of the neurodevelopmental disorders and psychiatric comorbidities, both the age 15-follow up studies featured the same written questionnaire to twins and parents with questions on asthma, allergic diseases, weight, and height. Finally, this questionnaire also asked for twins' consent to collection of growth data from their child and school health records. In this step there was some selection based on twins' sex and self-reported BMI at age 15. This aspect has been further discussed in one of the publications related to this thesis.[199]

Child health records are held at the child health care centres and include information on the child's weight, height, administered vaccinations and developmental checkups.[200] When the child starts school this practice is continued by the school health care system and documented in the school health record.[201] Child health records are most often stored in county council archives in the child's county of residence following the child's transition to school health care. Thus, during collection of **child growth data** child health records were requested from county archives, whereas school health records were requested from high schools.

Consent for growth data collection was provided when the twins were 15 years of age, and the actual record collection proceeded afterwards. This collection included twins born 1993-1996. The full growth data within CATSS-15 and CATSS-15/DOGSS includes n=3,172 twins.

Within this thesis, these cohorts were primarily the setting of *Study II.*

5.1.1.4 STOPPA

The Swedish Twin study on Prediction and Prevention of Asthma (STOPPA) includes 9- to 14-year-old twins specifically selected from CATSS-9/12 based on their asthma status. The total number of twins who participated in STOPPA was 752.

For STOPPA, the intention was to recruit a study population with three groups of approximately equal size. These were healthy concordant pairs (both twins in a pair had no history of asthma), asthma concordant pairs (both twins had a history of asthma), and asthma discordant pairs (one twin with and one without asthma). Each of the concordance groups included both DZ and MZ pairs.

A publication describing the full data collection in STOPPA in detail is available.[202] In brief, twins and parents were invited to one of seven clinical examination test sites (located in Stockholm, Gothenburg, Linköping, Lund, Lomma, Växjö, and Umeå). At the clinical examination, twins and parents each responded to written questionnaires. Twins also participated in clinical examinations (including lung function tests by spirometry and fractional exhaled nitric oxide) and provided biosamples. Blood, saliva, faecal and urine samples were all collected, some (faeces and urine) only in certain waves of the study. The blood sampling included two 4 ml EDTA tubes; one of these was stored at the KI Biobank for later DNA methylation analysis, and the other was sent to the clinical laboratory nearest to the test site for white blood cell count (WBC) measurements.

The STOPPA cohort was the setting of *Study IV.*

5.1.1.5 The Medical Birth Register (MBR)

The Medical Birth Register (MBR) started in 1973 and covers 98% of all births in Sweden. The register includes information regarding maternal health indicators from antenatal care visits (height, weight, some diseases, and smoking during pregnancy), as well as information registered at child's delivery (child's birth weight, gestational age, Apgar score, and mode of delivery, among other variables).[203] Within the work of this thesis, the MBR was the source of covariates for *Study I* and *Study II*, and also where multiple births were identified in *Study I.*

5.1.1.6 The National Patient Register (NPR)

The NPR covers all diagnoses given in in-patient care (from 1987 onwards) as well as those given in outpatient specialist clinics (from 2001 onwards). Visits to the emergency department in Swedish hospitals are not considered inpatient care unless the patient is later admitted.[204] The NPR does not cover diagnoses given in primary health care.

The diagnoses are recorded using the International Classification of Disease (ICD) system. In ICD version 10, diagnosis codes J45 and J46 are used for asthma; the corresponding code in the previous ICD version 9 is 493. Along with the diagnosis code, the dates of admittance, diagnosis, and discharge from the hospital are also recorded. Either of these can then be used as the date of diagnosis depending on one's preferences and the research question at hand. In many cases, the dates overlap.

Data from the NPR were used to construct outcome variables in *Study I, Study II,* and *Study III*.

5.1.1.7 The Swedish Prescribed Drug Register (SPDR)

The Swedish Prescribed Drug Register (SPDR) started on 1 July 2005. It contains information on all prescription medications dispensed to the Swedish population. Each observation in the register reveals the date of prescription, date of dispensing, and the Anatomic Therapeutic Chemical (ATC) code of the medication prescribed.[205] Most medications used to treat and manage asthma are sorted under ATC code R03 (obstructive lung diseases). The R03 group is further classified into the chemical subgroups.[206] The full list of medication dispenses used to construct the asthma outcomes within this thesis can be seen in *Table 1*.

An inherent weakness in the SPDR is the inability to distinguish between dispensed and actually ingested medications. However, if a person continues to return to the pharmacy and fills a similar prescription several times, it becomes more likely that it has also been used. Also, there can be questions regarding whether the medication is used for the disease intended to be studied – the validity and specificity of different medications may vary between diseases. For asthma, this point is of particular interest as there are other conditions with wheezing with which it may be confused. Therefore, there have been efforts to construct combinations of outcomes that are validated to better correspond to an actual asthma diagnosis.[207] These outcomes can then be used on their own or as a complement to asthma diagnosis from the National Patient Register.

Data from the SPDR were used to construct outcome variables in *Study I, Study II,* and *Study III*.

Table 1. *List of medications the dispensing of which (as registered in the SPDR) was used to construct the validated asthma medication outcomes. Brand names extracted from fass.se.*

ATC code	Group name	Active substances included	Sold in Sweden as…
R03AC	Selective beta-2-agonists	Salbutamol	Airomir, Airomir Autohaler, Airsalb, Buventol Easyhaler, Salbutamol Arrow, Ventilastin Novolizer, Ventoline, Ventoline Diskus, Ventoline Evohaler
		Terbutalin	Bricanyl, Bricanyl Turbuhaler
		Salmeterol	Serevent Diskus, Serevent Evohaler
		Formoterol	Formartris Novolizer, Oxis Turbuhaler
		Indacaterol	Onbrez Breezhaler
		Olodaterol	Striverdi Respimat
R03BA	Inhaled corticosteroids (ICS)	Beclomethasone	AeroBec, AeroBec Autohaler, Beclomet Easyhaler
		Budesonide	Budesonid Arrow, Budesonide Teva Pharma, Giona Easyhaler, Novopulmon Novolizer, Pulmicort, Pulmicort Turbuhaler
		Fluticasone	Fluticasone Cipla, Flutide Diskus, Flutide Evohaler
		Momentason	Asmanex Twisthaler
		Ciclesonide	Alvesco
R03AK	Fixed combinations of beta-2-agonist and ICS	Salmeterol + Fluticasone	Airflusal Forspiro, Salmeterol/Fluticasone Cipla, Seretide Diskus, Seretid Evohaler
		Formoterol + Budesonide	Bufomix Easyhaler, DuoResp Spiromax, Sybmicort Turbuhaler
		Formoterol + Beclomethasone	Innovair
		Vilanterol + Fluticasone furoate	Relvar Ellipta
		Formoterol + Fluticasone	Flutiform
R03DC	Leukotriene receptor antagonist (LRTA)	Montelukast	Montelukast, Singulair

5.2 STUDY POPULATIONS AND DESIGNS

5.2.1 Defining asthma

There are several different ways to identify asthma cases. The main difference between the clinical situation and research is that in the first, the examiner has direct access to the patient. In research, this is not always the case. Without the clinical examiner's possibilities to ask clarifying follow-up questions, perform tests first-hand, or request consultations from colleagues, the researcher needs to rely on the reports of others. The following section describes a few ways asthma may be defined under these circumstances.

5.2.1.1 Register-based methods of identifying asthma cases (Study I, II, and III)

In Sweden, the population-based clinical registers of medication dispensing (SPDR) and diagnoses (NPR) are highly useful as proxies for the medical researcher. Still, using these registers requires some qualified data mining. To make the most of the data, the sources should be used to complement each other.

A main issue with the SPDR is that the same medication may be prescribed for different diseases. Simply by looking at one filled prescription we may not be able to identify which. The beta-agonists prescribed may just as well be intended as temporary relief during a severe respiratory infection in an otherwise healthy individual as for someone with an asthma diagnosis.

Conversely, using a purely-register based definition of asthma diagnosis is limited by the fact that, as stated previously, the NPR only covers inpatient care. The asthma diagnoses recorded in this register will be the more severe cases; the less severe cases may be handled entirely in primary care.

And this is where the beauty of combining the data sources comes into play. The SPDR covers all prescriptions filled outside hospitals – also those prescribed by primary care physicians. Using both data sources thus allows us to capture some cases that would otherwise be missed.

To circumvent this problem of dispensed medication being misclassified, a validation study was performed by Örtqvist et al and published in 2013.[207] In the validation study, medical records of individuals fulfilling predefined asthma medication criteria were requested from primary care clinics and hospitals all over Sweden. The record entry from the day of prescription was reviewed to see whether the individual also fulfilled asthma diagnosis criteria. In this manner, two different medication outcomes were defined, which both had high positive predictive value (PPV) for clinically diagnosed asthma. The two asthma medication outcomes are defined in *Figure 5*. They were usually used in combination (i.e. fulfilling either of the two definitions was sufficient).

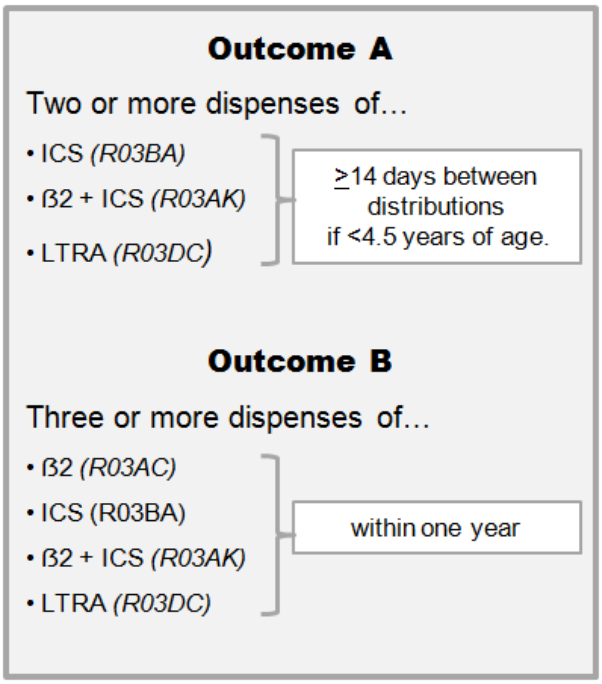

Figure 5 *Definitions of the validated asthma medication outcome combinations from Örtqvist et al.*[207]

ICS = inhaled corticosteroids ß2 + ICS = fixed combination of selective beta-2-agonists and ICS. LTRA = Leukotriene Receptor Antagonists. ß2 = Selective beta-2-agonist only.

The validation study also showed that on their own, the medication outcomes were more accurate after than before 4.5 years of age. Therefore, in *Study I* we also performed sensitivity analyses using a combined outcome where both asthma medication and a register diagnosis were required.

5.2.1.2 Questionnaire-based assessment (Study II, III, and IV)

Finally, one can obviously ask the study participants themselves whether they have asthma or not. Or, in the case of studies of children, one usually asks their parents. This was the case in *Study III* and *Study IV*, where parental telephone interview (*Study III*) or questionnaire (*Study IV*) was used as a source of information on child's asthma ever (*Study III*) or currently (*Study IV*). *Study II* featured questionnaires to both twins and parents, and although these did not always agree we opted to include as cases all twins for whom asthma had been reported, regardless of who reported it.

5.2.2 Overview of *Study I*

5.2.2.1 Study population

The study population in *Study I* was identified through the MBR and divided into two groups of children: one 'older' and one 'younger' group. The older group was born between 1 January 1993 and 1 June 2001. The younger group was born between 1 July 2005 and 31 December 2009.

Cut-offs in terms of birth dates were determined on the basis of ensuring coverage of the study participants in the SPDR. For the older group, we intended to ensure that each study participant had at least 5 years of data available before their 18th birthday. In the younger group, we aimed to include all individuals who had data in the SPDR from birth, and who were then covered in that register for at least one year onwards (so as to allow them to fulfil medication outcome B during this time).

5.2.2.2 Study design

Study I is a population-based register study of all twin and singleton children born in Sweden between the aforementioned dates. Due to the different setups in the two cohorts, different analysis methods were used.

In the younger cohort, with complete coverage in both outcome data sources from birth, we were able to detect the first time they experienced any of the outcomes. Knowing the date of this event enabled us to use survival analysis (Cox Proportional Hazards Regression) in this group.

In the older cohort, the two outcome data sources covered slightly different points in time, and most of the participants were not covered in the SPDR from birth (the youngest participants in the older cohort were just above 4 years of age when it started). Therefore,

analyses based on the prevalence of asthma across all the covered time were deemed preferable. Logistic regression was used in this group.

5.2.2.3 Exposure and outcome definitions

Twinship was the exposure and childhood asthma (as defined by medication, diagnosis, or both) was the outcome. In the older group, information on dispensed medication and asthma diagnoses before 5 or after 18 years of age was not included.

5.2.2.4 Notes on measurement methods in Study I

Multiple birth status in the MBR

The MBR contains information on whether a birth is singleton or multiple, but as triplet and higher order multiple births are also included in this variable these had to be manually counted and excluded. This was done by summing the number of births to the same mother within +/-1 days of the 'index' child's birth (to allow for the possibility of multiple birth siblings being born on different sides of midnight). If this number equalled 3 or higher, the children were not eligible for the study and all observations in this higher order multiple group were dropped.

Mode of delivery

We considered this an interesting potential confounding factor for *Study I* as twins are more often delivered by caesarean section, and there has been a debate in the literature regarding whether C-sections increase the risk of asthma.[208] As elective caesarean sections are on the rise for various reasons,[209] they are becoming more common also for singleton deliveries.

Using the same categorisation as in a previous Swedish study that was based on similar data,[208] we were able to some degree to distinguish between elective (planned) and emergency caesarean sections – which is of interest as previous studies have indicated that these different indications for the same delivery method may be associated with different childhood asthma risk patterns.

5.2.3 Overview of *Study II*

5.2.3.1 Study population

For *Study II*, we selected all study participants in either CATSS-15 or CATSS-15/DOGSS for whom childhood growth (weight or height) data from 0-3 years of age had been successfully collected. This resulted in a study population including n=2,874 twins.

5.2.3.2 Study design

The exposure data were longitudinally recorded at child health care centres prior to the study initiation. The individual CATSS cohorts can be described as retrospective studies, but the fact that there are several follow-ups at various ages adds a longitudinal dimension. Additionally, the CATSS study participants have been linked to several population-based

registers, including the NPR and SPDR – the coverage of which sometimes exceed the time period covered by the questionnaires themselves.

5.2.3.3 Exposure and outcome definitions

For weight and height separately, childhood growth parameters were modelled using SuperImposition by Translation and Rotation (SITAR; further explained in *section 5.3.6, Growth modelling*).[210] Each SITAR model parameter (a, size; b, tempo; and c, velocity) for each growth dimension was used as its own separate exposure in a logistic regression model.

As the study used data from CATSS-15 and CATSS-15/DOGSS, data from CATSS-9/12 were also available for many of the study participants. The asthma outcome was constructed as asthma according to reports by either the twin or a parent in any of the STR interviews/questionnaires, or as identified from the NPR or SPDR before age 18. Finally, in an additional analysis, this asthma was divided into disease that first occurred before or after 3 years of age.

5.2.3.4 Notes on measurement methods in Study II

Childhood growth data

In *Study II*, weight and height measurements from birth up to 3 years of age were recorded in childhood health records: routinely kept medical records that are established for each child once they enter child health care. Initially, the child is measured lying down on a scale. At around two years of age, the child starts to be measured standing up.[211] Once measured, the age at or date of measurement is noted in the child health record. In the present day, when most records are digitally kept, the growth curve for the child is then plotted automatically. In the 1990s, when the participants of *Study II* were born, however, plotting was performed manually by the school health nurse. We used the exact date of measurement to calculate the age whenever possible, but sometimes had to substitute a missing value by reading off the child's growth curve. This of course introduces a potential for human error; however, as those extracting data from the child health records did not have access to information on the asthma outcome, these reading errors should be spread out evenly between the groups (non-differential misclassification).

5.2.4 Overview of *Study III*

5.2.4.1 Study population

Study III included all twins who were responders in CATSS-9/12 as of October 2014, n=25,306. A subset of these twins (n=10,075) had been genotyped (see details on page 28) and data on 16 SNPs previously associated with childhood asthma were available.

5.2.4.2 Study design

Study III combined two main study designs: a quantitative genetic study of the heritability of asthma and related phenotypes, and a genetic association study of 16 SNPs and these same phenotypes.

5.2.4.3 Exposure and outcome definitions

Study III deals with phenotypes (asthma and commonly comorbid allergic diseases: these were hay fever, atopic eczema, and food allergy) and genetic exposures (either known single-nucleotide polymorphisms, SNPs, or the more abstract heritability concept).

The phenotypes were defined as follows:

1) **asthma** was defined based on a combination of different data sources; these were the CATSS-9/12 telephone interview, the SPDR and the NPR. Any asthma was defined as having asthma according to any of these data sources. Each data source was also used to construct its own outcome.

2) **wheezing** was from CATSS-9/12, and is here to be seen primarily as a complement to asthma. Due to the design of the telephone interview, the question regarding wheezing was only asked when the prior question (about asthma ever) had been answered in the negative. Thus, this was only such wheezing that had not already been classified as or later developed into asthma by age 9 or 12. Likely, this is a particular phenotype (including the so-called early transient wheezers) but which differs somewhat from many other studies where wheezing has been independently assessed.

Based on additional information regarding symptom duration, wheezing was further classified into that which occurred ever or lasted until after three years of age – a distinction intended to at least in part distill out the early transient wheezers.

3) **hay fever**, **food allergy**, and **atopic eczema** were based on parental reports in CATSS-9/12.

5.2.4.4 Notes on measurement methods in Study III

Zygosity

The most important purpose of zygosity testing is to distinguish between genetically identical (MZ) and same-sexed non-identical (DZ) twins. Opposite-sexed twin pairs are always DZ. There are two primary sources of zygosity assessment in CATSS: a question-based algorithm based on perceived physical similarity between the twins[195] and DNA testing of both twins using an array featuring 46 predetermined SNPs.[197] Each method has been validated with >95% accuracy. When available, DNA-based zygosity assessment took precedence over that based solely on the algorithm. In the *Study III* study population, 58% of twins (n=14,608) had their zygosity assessed by the algorithm, and 41% (n=10,398) by DNA.

Genotyping

For the purposes of this study, 18 SNPs were selected to be genotyped. However, only 16 of these SNPs were successfully genotyped. A full list of the genotyped SNPs is available in *Table 2*, below. The two SNPs which could not be genotyped were both located in the HLA region on Chromosome 9, and were previously associated with asthma[146] or total serum IgE levels.[212] Genotyping was performed at KBioscience in Hoddesdon, Herts, UK.

Table 2 *List of SNPs that were selected for genotyping in Study III, by chromosome (Chr), closest gene, position in relation to said gene, SNP name, minor allele, and the discovery study from which the SNP was identified.*

Chr	Gene	Location	SNP	Minor allele	Source
1	IL6R	Intron	rs4129267	T	Ferreira et al 2011[148]
2	IL1RL1	Intron	rs3771180	T	Torgerson et al 2011[147]
2	IL18R1	Intron	rs3771166	A	Moffatt et al 2010[146]
5	TSLP	Nearby region	rs1837253	T	Torgerson et al 2011[147]
5	SLC22A5	Intron	rs2073643	T	Moffatt et al 2010[146]
5	IL13	Intron	rs1295686	T	Moffatt et al 2010[146]
6	HLA-DQB1	Nearby region	rs9273349	Genotyping failed	Moffatt et al 2010[146]
6	HLA-DQB1	Nearby region	rs9469220	Genotyping failed	Levin et al 2013[212]
9	IL33	Nearby region	rs1342326	C	Moffatt et al 2010[146]
9	IL33	Nearby region	rs2381416	C	Torgerson et al 2011[147]
15	RORA	Intron	rs11071559	T	Moffatt et al 2010[146]
15	SMAD3	Intron	rs744910	A	Moffatt et al 2010[146]
17	Z2BP2	Intron	rs12936231	C	Verlaan et al 2009[151]
17	GSDMB	Exon	rs2305480	A	Moffatt et al 2010[146]
17	GSDMB	Intron	rs11078927	T	Torgerson et al 2011[147]
17	GSDMB	Intron	rs7216389	T	Moffatt et al 2010[146]
17	GSDMA	Exon	rs3894194	A	Moffatt et al 2010[146]
22	IL2RB	Intron	rs2284033	A	Moffatt et al 2010[146]

Although only 10,075 twins had been genotyped themselves, due to the fact that some of these were MZ the SNP values for the same genotypes in their non-genotyped MZ co-twins could be imputed. Following this imputation, 12,388 twins had genotype data.

As retaining both twins from an MZ twin pair in genetic association analysis would have inflated the estimates, we randomly selected one MZ twin from each pair to keep in the statistical analysis. This could be either the twin whose DNA was originally analysed, or the co-twin whose genotype was imputed based on the first twin's sample.

5.2.5 Overview of *Study IV*

5.2.5.1 Study population

The study population in *Study IV* consisted of all twins in STOPPA who had contributed at least one whole blood sample that could be used for DNA methylation analysis. The total number was 708 twins (94% of the total study population in STOPPA).

5.2.5.2 Study design

Study IV is a cross-sectional cohort study featuring detailed clinical examination focused on childhood asthma.

5.2.5.3 Exposure and outcome definitions

DNA methylation at ~485,000 CpG sites was measured using the Illumina HumanMethylation450 BeadChip (Illumina 450k).[137] For the primary analyses, current, parent-reported asthma at the time of the clinical examination was used. This meant that some twin pairs were re-classified following the selection to the STOPPA cohort, as they had received an asthma diagnosis between the time of the participation in CATSS-9/12 and the later clinical examination in STOPPA.

5.2.5.4 Notes on measurement methods in Study IV

DNA collection, storage, and extraction

Blood samples to be used for later DNA methylation analysis were collected in 4 ml EDTA test tubes at the STOPPA test centres. To ensure that these samples were acquired from as many twins as possible, they were usually the first samples to be taken. After clinical examination, samples were stored at -20 degrees Celsius during transport to their long-term storage location (usually arriving within the next two days), after which they were kept in freezers at -80 degrees Celsius until the point in time at which they were analysed. All DNA was extracted at the KI Biobank using the Chemagic STAR DNA Blood 400 kit.

DNA methylation analysis and plate layout

DNA methylation analysis was performed at the Mutation Analysis Facility (MAF) at Karolinska Institutet Campus Huddinge. Sample plates contained 96 wells spread over 8 chips featuring 12 samples each in a predetermined **plate layout**. This plate layout was designed to minimise the risk of results being affected by so-called batch effects, which are systematic errors that may arise in this type of laboratory analysis.[213] Within each 96-well-plate, 5 control samples were included. These control samples included one test sample of

known methylation pattern (ZYMO-1331), one sample with high methylation (ZYMO-high), one sample with low methylation (ZYMO-low), and two samples that were copies of real test subject samples from the same plate. Apart from the fact that none of the control samples should occupy the top right sample well, they could be randomly allocated anywhere on the plate. The remaining 91 wells could be used for study subject samples. The position of all test subject samples was randomised across plates and chips to the highest degree possible. However, as our core comparison in the study is the within-pair analysis in (discordant) twin pairs,[214] we elected to keep twin pairs' samples together within the same chip.

Because not all twin pairs both contributed blood to the study even though they otherwise took part in the clinical examination, there were 32 'single' twins whose samples were used to balance out uneven numbers (resulting from the fact that there was an uneven number of control samples to be included). A figure illustrating one of the potential results of this plate layout system (specifically, the plate layout of plate 7) is shown in *Figure 6*.

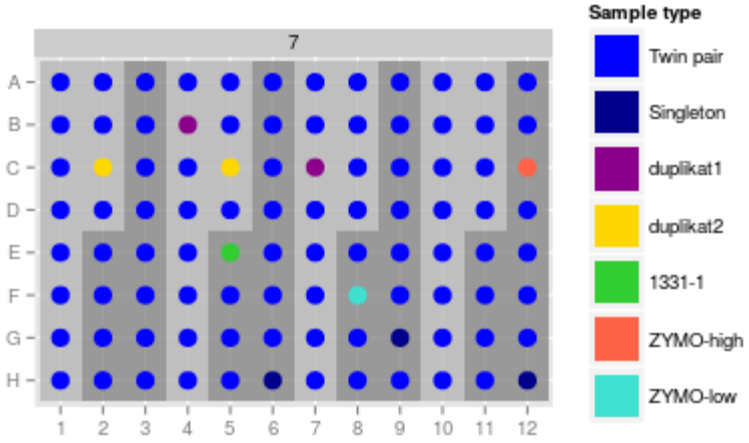

Figure 6 *Plate layout of plate 7 in the DNA methylation analysis in STOPPA.*

White blood cell (WBC) counts

White blood cells (WBC), or leukocytes, are one of the primary cell types circulating in the blood. They are also the only blood cells to carry DNA. Thus, in any given sample, the DNA methylation patterns represented will primarily be those of the leukocytes.

In STOPPA, WBC counts were analysed at local laboratories close to the clinical test centres. The second 4 ml EDTA sample was used for this purpose.[202] The WBC counts are made up of a total leukocyte count, as well as the number of cells of each subtype. The primary subtypes were the granulocytes (neutrophils, eosinophils, and basophils) and lymphocytes (including T- and B-cells). Granulocytes were further divided into neutrophils, eosinophils, and basophils.[33] Basophils are very few in number. Neutrophils and eosinophils, on the other hand, are numerous, and levels may vary with disease, including asthma.[215] When this occurs to such a degree that the cell type becomes over- or under-represented in a blood sample, it could affect the DNA methylation pattern in the sample.[127] This will be referred to as **confounding by cell type**, i.e. that differential DNA methylation detected in the sample is that of a specific cell type and not necessarily something specific to the disease itself.

5.3 STATISTICAL METHODS

5.3.1 The nature of statistics

Briefly, statistical analysis in epidemiology is intended to describe the relationship between an exposure (something an individual is, does, or has been exposed to – often referred to as a potential risk or protective factor) and an outcome (usually a trait or disease, such as asthma). This is done by collecting a set of data on both exposure and outcome in a number of individuals, and thereafter making calculations based on these data or imposing a statistical model on them. Then, conclusions are made based on the patterns of associations found within the results of these calculations.

The choice of model or calculation performed depends on several factors; whether the exposure and/or outcome is continuous or categorical, if they vary over time, if some individuals do not have all the data, and how the variable is distributed within the population all come into play. A common basic prerequisite for many statistical models is that the observations are completely independent from each other. This is violated in twin and family studies. In these, this problem needs to be circumvented by somehow accounting for the clustering of observations within specified groups (i.e. families). On the other hand, other types of models rely on these correlations to draw specific conclusions about genetic and environmental components.

The following section introduces the statistical models and analytical methods applied within the work of this thesis. First, however – a note on association versus causation.

5.3.2 Association is not causation – the benefits of Directed Acyclic Graphs

Directed acyclic graphs – or DAGs – are a helpful tool to visualise the supposed relationships between an exposure, an outcome, and various other factors which may be connected to these variables of interest.[216] The primary purpose of this visualisation is two-fold: to increase the clarity of communication regarding the exposure–outcome association that is in fact under study, and to ensure that no extraneous factors irrelevant to the relationship at the heart of the question are allowed to intrude in the model. When such intrusion happens, it may be harmful not only because it risks weakening the statistical power, but because adjusting for some of these factors (so-called colliders – more on these later) may introduce bias.

DAGs (of which *Figure 7* presents an example, which serves to illustrate the concepts outlined in the current and following section) imply some assumptions and basic requirements. The arrows imply not only an association, but also state the direction: thus saying that it is variable A that affects variable Y and not the other way around. Unless the researcher has a clear idea about the explanation for the potential association between A and Y, this direction may actually be somewhat difficult to determine.

In addition to deciding on the direction of an arrow, one also needs to decide whether it's supposed to be there in the first place. Omitting to draw an arrow between two factors in a DAG (such as between X and L in the example) is a strong statement, as it means the person

who constructed the DAG has strong reasons to believe there is in fact no direct connection between these two factors.

5.3.3 Confounders, colliders, and mediators

The construction and display of a DAG may be a helpful exercise in itself. But the next step is when one can unleash its true power: that of discerning which factors acting in the vicinity of an exposure and outcome are the potential **confounders**[217] that should be adjusted for, **mediators** that may account for part of the effect between the exposure and the outcome, or **colliders** that if controlled for may lead to biased results.[218]

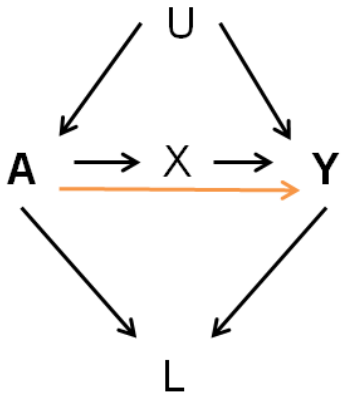

Figure 7 *An example DAG. A is the exposure, Y is the outcome, X is a mediator between A and Y (accounting for some of the total effect), U is a confounder, and L is a collider. The orange arrow A->Y represents the direct effect of A on Y. The total effect of A on Y goes through the pathways A -> Y and A -> X -> Y.*

An association between an exposure and an outcome, provided that it exists, is sometimes made up of a sum of different effects. There is always a **total effect**: this is the association seen if one estimates the association between A and Y without accounting for any potential confounding whatsoever. Then there is the **direct effect**: the part of the association that is purely due to A and Y, without going through any factors in between. And finally, there is the potential for **mediated effects**: that A acts on Y through another factor X. The total effect of A on Y is the sum of the direct and any mediated effects.

If the direct effect is the association of interest, any potential **backdoor pathways** between A and Y must be closed.[219] In this example, this would be accomplished by conditioning the analysis on X (the mediator) and U (the confounder). However, **umeasurable or unknown** confounding may remain.

5.3.4 Logistic regression (*Study I*, *II*, and *III*)

Used in *Study I*, *Study II*, and *Study III*, logistic regression models were perhaps the most prevalent statistical analysis methods applied within this thesis. They are a good fit in many studies and for many types of data (so long as the outcome under study is categorical and binary, for instance asthma yes (=1) or no (=0)). Another assumption of the logistic

regression model is that the observations within the data set should be unrelated to each other – this includes the exposure variables included in the model. If not, clustering or interaction may need to be considered. This was the case with sibling constellations (*Study I*) as well as twin pairs (*Study II* and *Study III*).

The effect estimate of logistic regression models is the odds ratio (OR), usually presented together with a 95% confidence interval (CI). Clustering within pairs or groups is often accounted for by calculating robust standard errors using a clustered sandwich estimator, which in turn widens the confidence intervals.[220]

5.3.5 Cox Proportional Hazards Regression (*Study I*)

Cox Proportional Hazards Regression is a form of survival analysis that was used in *Study I*. Briefly, survival analysis presents the effect estimate as a hazard rate: the number of events occurring at the time t, in which the denominator is the number of remaining individuals who have yet to experience the event. In *Study I*, attained age was the underlying **time scale.** The possible **endpoints** of the study were: attaining the outcome (for asthma medication, the date of the first dispensing of any prescription in a medication combination that fulfilled the outcome was used as the event date), death, and migration (i.e. leaving Sweden and thus no longer being covered by the population-based registers).

The **proportional hazards assumption** states that the relative difference (hazard ratio) should be constant between exposed and unexposed groups over time.[221] In each case of analysis, it should be examined whether this is true. Such examination can be performed either by inspecting a curve and performing a visual assessment, or by testing in post-estimation following the execution of a model. When the proportional hazards test fails, separate effect estimates may need to be presented for different age groups. This was the case in *Study I*, where the results in the younger cohort were presented over different time intervals (0-1.5 and 1.5-5.9 years of age).

5.3.6 Growth modelling (*Study II*)

For *Study II*, we used the SITAR – SuperImposition by Translation and Rotation – growth model. SITAR produces three parameters for each individual: a (size), b (tempo) and c (velocity).[210] Each of these describes individual growth of each child as it compared to the average curve in the whole population, and can be applied to either weight or height change.[222]

The basic structure of a fitted population growth curve resulting from this model is shown in *Figure 8.* The x axis is usually a time scale: in this study, we had access to the exact date of most measurements, and could therefore calculate the age in years down to decimals. The y axis is the growth measurement, in this example: weight in kilograms (kgs).

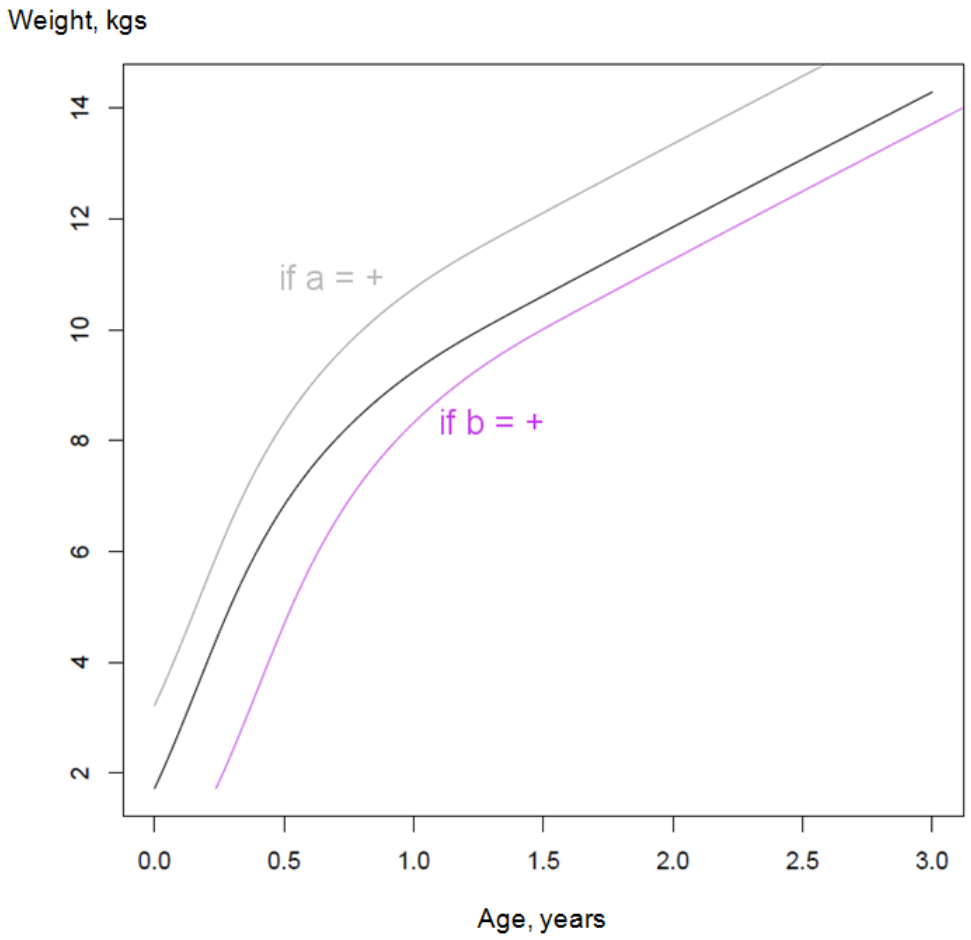

Figure 8 *Fitted population growth curve from a SITAR model (black line) in Study II. The light grey and magenta lines illustrate the potential effects of positive shifts in the a or b parameter.*

The interpretation of the individual growth parameters in this context is as follows.

a (size) represents the shift of the individual's growth curve on the y axis; in other words, how large the individual is compared to the general study population. A positive value of a represents a larger size than the average, whereas a negative value is a smaller size. The unit of measurement of the a parameter is the same as the y axis: here, it would be kgs.

b (tempo) describes the timing of the maximum growth velocity; whether an individual has their 'growth spurt' (provided a classical such exists within the time window of interest – if not, it will simply be point at which the most rapid growth within the interval takes place) before (negative value of b) or after (positive value of b) the population average. It is thus a shift of the growth curve along the x axis.

Finally, the ***c*** (velocity) parameter is the natural log of a multiplier of the slope of the growth curve compared to the population average. Those with a positive c have had steeper growth than the population average.

By nature of the character of growth during different points of time in childhood, the SITAR parameters are correlated. Therefore, once estimated, the growth parameters were included as

exposures in their own separate logistic regression model (i.e. one model for a, b, and c, for weight and height, respectively).

In *Study II*, we selected the time window 0-3 years of age, and included all available measurements of growth for each child during this period. In total, 2,874 twins had data during this period. As infant growth patterns vary slightly between males and females,[223] we estimated separate average curves for each group. In addition, each of the resulting parameters was inspected in terms of the mean and standard deviation for differences between these groups. Although this difference was not statistically significant, we still opted to standardise the parameters by sex.

5.3.7 Structural Equation Models – Twin modelling (*Study III*)

Structural equation models are often applied in the context of twin modelling. Based on correlations of a phenotype within twin pairs for which a certain predetermined genetic correlation is assumed (0.5 for DZ and 1.0 for MZ twins), the remaining proportions and sources of variation can be calculated.[187] Heritability is one such factor, usually termed 'additive genetics', or A, in classic structural equation twin models.

Univariate heritability models focus on one trait or phenotype, and estimate the degree to which components A (additive genetics), C (shared environment) or D (genetic dominance), and E (non-shared environment) contribute to the total variance of this phenotype within the population.[224] The sum of these contributing factors is always 1, i.e. the whole. Additionally, bivariate heritability models can be constructed to look at the shared heritability of two different phenotypes.[225] Within the scope of this thesis, however, all the heritability models that were constructed were univariate.

The assumptions underlying structural equation twin models include:

1) The equal environments assumption: stating that we assume that MZ and DZ twins share their environment to an equal degree within the pairs.

2) There is no gene-environment interaction

and

3) There is no assortative mating

Of these, the non-assortative mating assumption has been challenged recently. One recent study has shown that when it comes to psychiatric diseases, this assumption likely does not hold.[226] However, as we have yet to see such evidence for asthma, in this case we have elected to assume that the assumption is mostly correct.

Finally, twin models are formulated as either ACE, ADE, or AE, depending on the correlations within different zygosities (ADE was selected when the tetrachoric correlation in MZ pairs was more than twice that of the DZ-ss twins) and if a simpler model would suffice (AE was used in place of A(C/D)E if a likelihood ratio test (LRT) comparing the more

complex to the simpler model revealed that the extra parameter did not significantly improve the model fit). The software used to construct the twin models and perform these calculations was OpenMx.[227]

5.3.7.1 When the importance of genetics varies by sex

Typically, only same-sexed DZ and MZ pairs are included in twin modelling. This was also the case in the main analysis in *Study III*. The reason for this is a wish to circumvent the risk that potential confounding by sex would somehow influence the associations. However, the prevalence of asthma is known to differ by sex, with the age group under study here being of particular interest (as it is just before the 'puberty switch' – at which the prevalence of asthma changes from being higher in males to instead being higher in females[228]).

In addition to the basic features of univariate twin models, opposite-sexed DZ twin pairs can provide extra information which cannot be attained from using only the same-sexed DZ pairs. In this case, sex limitation models can be specified.[229,230] These constitute varieties of structural equation models that allow for either

1) A factor rA which is a multiplier of sex (qualitative difference – this allows for different genes to be of importance for different sexes)

or

2) making separate estimates of each variance component by sex (quantitative difference – this implies we believe they are the same genes, but their importance differs).

To determine which of these – if any – was the case, the models were fitted in different combinations to the data, after which the fit of these models were compared via another LRT.

5.3.8 Differential methylation and limma models

Prior to the differential methylation analyses in *Study IV*, CpG probes were filtered based on intensity, proximity to a SNP, and location on sex chromosomes. The final number of CpGs retained after this process had been completed was 485,117 (94% of the original ~485,000).

After CpG cleaning and data normalisation using the dasen method,[231] we analysed the DNA methylation data using Linear Models for Microarray Data (limma),[232] as incorporated within the R package RnBeads.[233] Briefly, limma models are linear models specifically designed to handle data from microarrays such as the Illumina 450k Beadchip. The measure of DNA methylation at each CpG site going into this model is the M-value: the log2 ratio of methylated vs. unmethylated probes at each site. A related estimate, the beta value, is defined as the ratio of the intensity of the methylated to the sum of the unmethylated and the methylated probes.[234]

Differences in M value between cases and controls were calculated to identify CpG sites with significant differences in methylation. For these, absolute values of the average methylation

in each group were also presented. P-values obtained using this method were presented both in unadjusted and false-discovery-rate-adjusted (FDR-adjusted) form.

5.3.8.1 Confounding by cell type

We aimed to correct for the composition of each study participant's WBC by adjusting the limma models for one of two ratios: the neutrophils to leukocyte ratio (NR) or the eosinophil to leukocyte ratio (ER). When included as in the models, the ratios were treated as continuous variables. In the manuscript, we referred to this adjustment as correcting for **confounding by cell type**. However, it should be noted that we adjusted for each of these cell type ratios independently. This was done because they are correlated to each other, but also because they might capture slightly different things.

5.3.8.2 Confounding by genetics

The main method of adjusting limma models for confounding by genetics was by performing analyses within asthma-discordant twin pairs. For further explanation of the rationale behind this method, see below.

5.3.9 Within-pair and co-twin control analyses

Within-pair and co-twin control analyses are essentially part of the same concept, but the terms are used to imply slightly different applications of the underlying principle.
Briefly, the point of both analysis methods is to identify the twin pairs who are discordant for both the exposure and the outcome, and then to study the exposure–outcome association specifically within these pairs. This 'model', if one would like to refer to it as such, is then naturally controlled for all factors that are shared within the twin pair. For instance, the twins will have both been exposed to potential maternal smoking during pregnancy – this shared environmental factor can no longer confound any association remaining within the pair. In the case of MZ twins, the shared factors include their entire set of potentially segregating genes; which are only shared to on average 50% between DZ twin pairs.[235]
Within-pair analysis, for instance within asthma-discordant monozygotic twin pairs (*Study IV*), is thus a very elegant model allowing for interesting conclusions regarding what may still remain to have made these twins different in terms of disease.[214] These within-pair comparisons are often carried out in parallel within several discordant pairs of the same cohort, to attempt to identify differences that are shared between several twin pairs.
Co-twin control analysis, as the term is most commonly used, involves making these within-pair analyses both within MZ and DZ twin pairs. The strength of any potential remaining association is compared between the zygosity groups. As comparisons within MZ twin pairs allow for additional controlling for shared genetic factors, an association that was visible in DZ but disappears in MZ twin pairs may have been partially explained by the genetic variation that has now been completely controlled for. Because this method assumes that the underlying zygosity assessment is reliable, twin pairs with unknown zygosity should not be included.

6. ETHICAL CONSIDERATIONS

6.1 RISKS FOR THE STUDY PARTICIPANTS

Medical research involving human subjects carries with it special challenges and considerations no matter the topic of said research. The potential benefit of the research always needs to be balanced against the risk to which the study participants are subjected. Withdrawal of blood and other biological samples causes discomfort and may be an unpleasant experience for some study participants. Efforts to minimise this discomfort included providing the study participants with information beforehand, as well as providing patches with local anaesthesia cream for application prior to the clinical examination in STOPPA (*Study IV*).

As this clinical examination included some blood and lung function tests which could reveal current health issues, the results of these tests were reviewed by a specialist in paediatric medicine. Twins whose white blood cell counts lay outside of the normal range or whose spirometry results suggested underlying pathology were contacted following the clinical examination and encouraged to get in touch with their general practitioner.

6.2 INFORMED CONSENT

It is also important to remember that informed consent is a cornerstone of all medical research. The right to have research to which one is asked to participate explained in understandable terms, as well as the right to decline participation in said research, is governed by international directives[236] as well as Swedish law.[237] In Sweden, particular care is also taken to emphasise the considerations needed to be taken for medical research concerning minors.

In all of the studies included in this thesis, the study participants were younger than 18 years at the time of participation in the research. The approaches taken to consider this were slightly different between studies because the age of the study participants varied. One recommendation based in Swedish law states that an individual who is between 15 and 18 years of age should be allowed to make their own decision regarding their participation, provided that they can be expected to understand what the study entails on their behalf. Therefore, in *Study II*, twins' own consent was requested for growth data collection, whereas in *Study IV*, consent was collected from both twins and parents.

6.3 PRIVACY AND INTEGRITY

On the other hand, due to the extensive nature of the information collected in epidemiology, particularly in register-based and genetic studies, a larger concern is that of potential violations of integrity. Great care is taken to protect the privacy of study participants. In all research and analysis data sets, names, personal identification numbers and other pieces of information that can be used to identify a study participant have been removed, replaced with an anonymised study participant number.

6.4 DATA SECURITY

While handling detailed data including information on genetic and epigenetic variation, data security is of particular importance. Therefore, in addition to ensuring the integrity of individual study participants by anonymising the data, the data themselves are protected by storage on secure servers. For each project included in this thesis, data has been stored in the most restrictive way possible, granting access only to those individuals who are involved in the project in question.

There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after. — J.R.R. Tolkien, The Hobbit

7. MAIN RESULTS AND DISCUSSION

7.1 TWINS AND THE OUTCOME – DIVING DEEPER INTO *STUDY I*

Study I took place within population-based registers. The total number of study participants in the two cohorts was 1,247,952 – children born in Sweden over the course of 13.5 years. These were divided into two different cohorts: the older cohort (born 1993-2001) and the younger cohort (born 2005-2009).[238] We found slightly different results in each of these cohorts. Briefly, there was an increased risk of asthma for twins compared to singletons in crude analyses in the younger cohort – but not in the older. Adjusted models (covariates were gestational age, birth weight, Apgar score, mode of delivery, presence of older siblings, and parental education) showed lower risk of asthma in twins in both age groups.

Some studies of asthma in twins (or other multiple births) versus singletons were published before ours.[64,65,68,239-243] Together, these spanned a variety of ages, starting from birth [65,68,240] and going well into adulthood.[239,242] *Figure 9* shows available effect estimates (OR, HR, or RR) from these studies as well as our own.

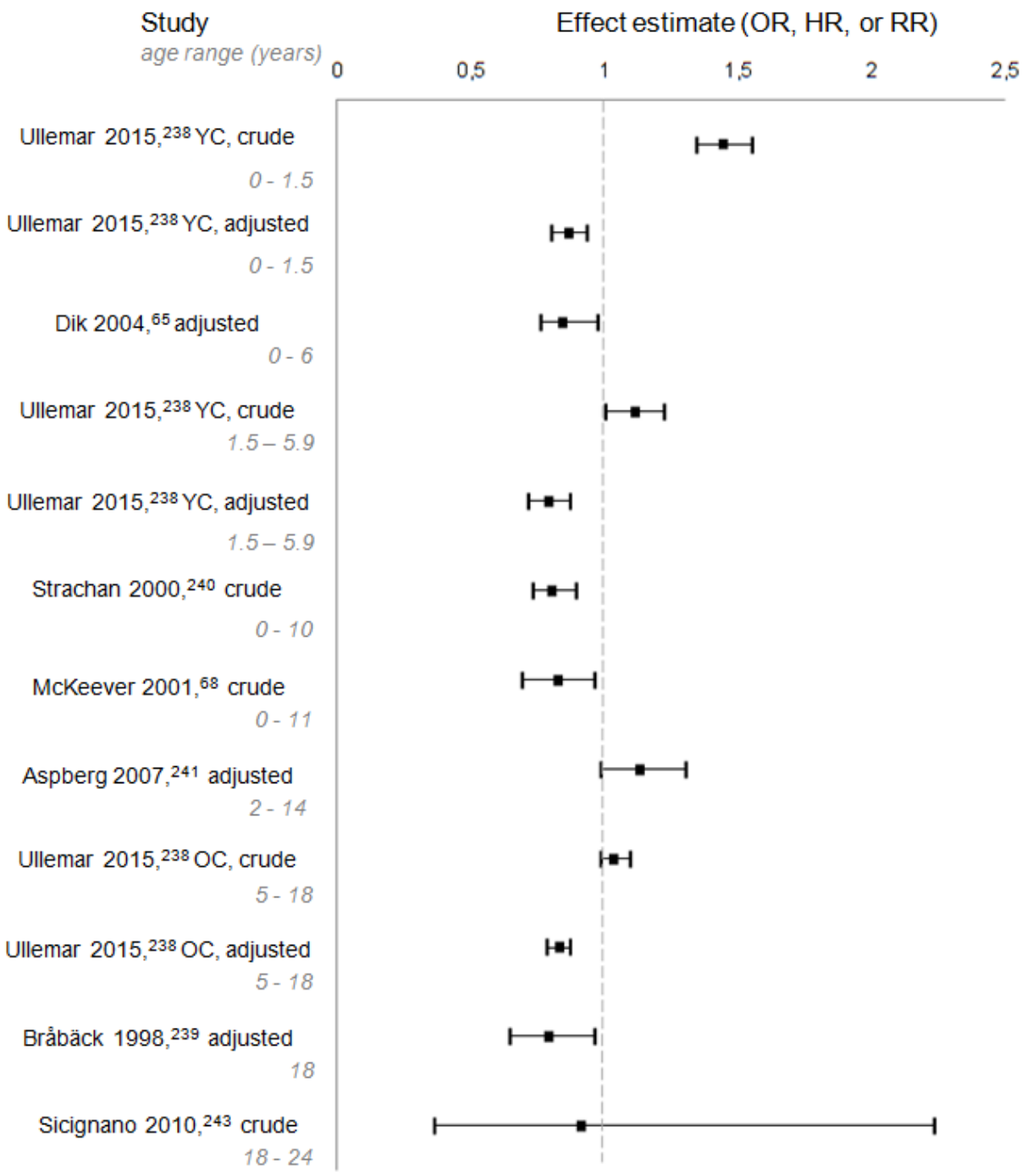

Figure 9 *Findings from available studies of twinship and asthma reporting effect estimates as either Odds Ratios (OR), Hazard Ratios (HR), or Relative Risks (RR). Estimates are presented roughly according to the age range of the participants (from youngest to oldest) and whether the estimate is crude or adjusted. YC = Younger cohort of Study I. OC = Older cohort of Study I.*

As seen here, there is no clearly discernible pattern within this collection of results. Additionally, the many facets of heterogeneity between the studies (including different statistical models) likely preclude direct comparisons of the estimates. Still, it is worth noting that the younger cohort (YC) in *Study I* is the only one in which there was a statistically significant higher risk of asthma in twins in crude analyses.

7.1.1 A true difference by age – or something else?

As our intention in defining the cohorts in this way was to be able to discern differences related to age, as well as to use different statistical models and effect measures, it is tempting to allow the explanation to remain centred on these factors. But what if it is something else?

Children age, yes – but the times change, too. One thing that has changed over the years is the proportion of children delivered by caesarean section.

Figure 10 (below) shows the proportions delivered by caesarean section (elective or emergency) in the twin vs. singleton group. The x axis displays the birth years (note the slight jump between 2001 and 2005: this was the gap between the two cohorts).

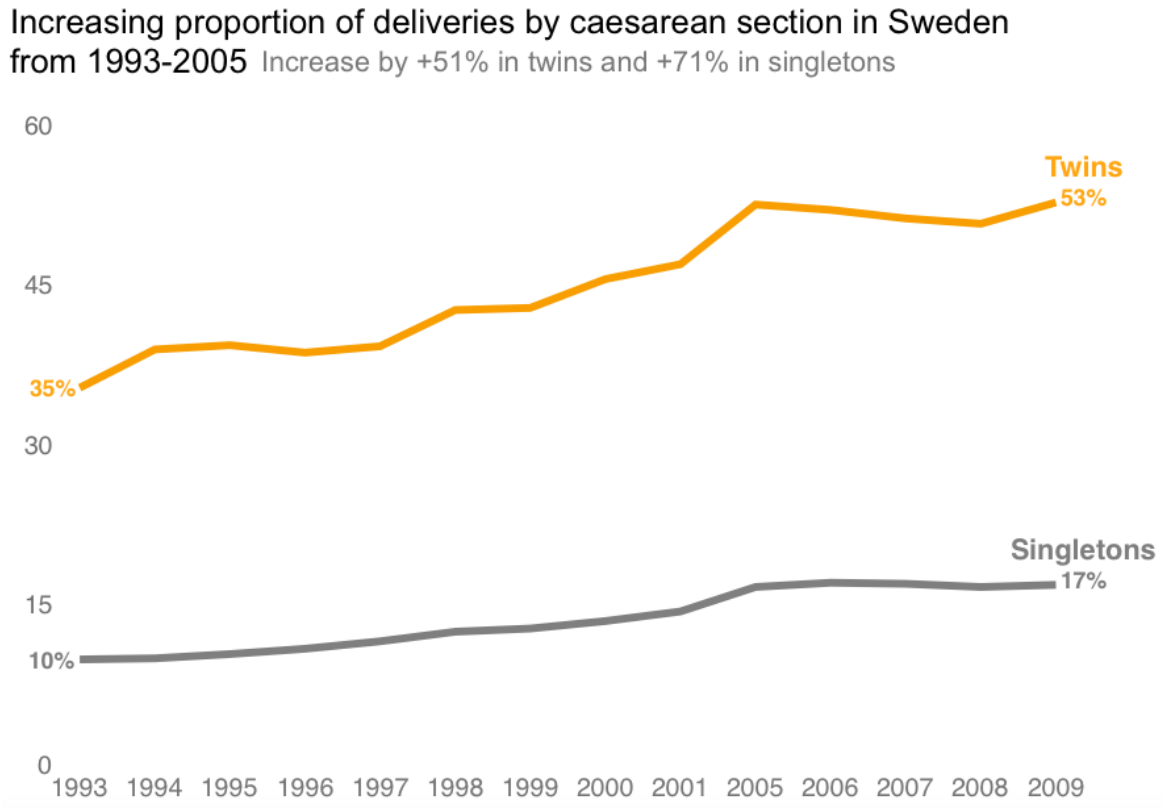

Figure 10 *Proportion of deliveries (%,on the y axis) by caesarean section in Sweden during the time period (calendar years, on the x axis) when the participants in Study I were born.*

By 2009, more than half of all twins were delivered by caesarean section (CS), compared to just over a third in 1993. However, the relative increase in this proportion within the singleton group during the same time period was actually larger (from 10 to 17%, a 71% increase). These figures reflect changed practices within this time period – particularly between 1993 and 2001.

While changed practices aiming toward the earlier delivery of twins could be expected to lead to an increase in elective CS in the twin group as a whole (which it did, by 29%), but there was also an increase in emergency (i.e. non-planned) CS during the time period. Indeed, this increase was larger: from 17 to 30%, a 76% increase. This change is shown over time in *Figure 11* (again, note the skip on the x axis due to the time gap between the cohorts).

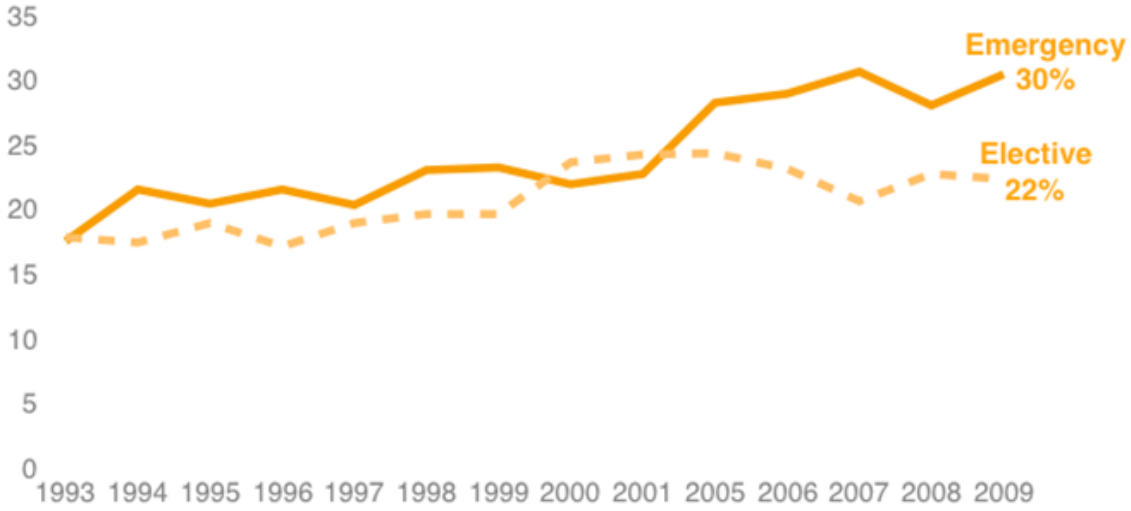

Figure 11 *Proportion of twin deliveries (%, on the y axis) by emergency or elective caesarean section in the Study I population – calendar years 1993 to 2009 (x axis). After each starting out at 17% of twin deliveries in 1993, proportions of both emergency and elective caesarean section increased by 2009.*

Among singletons, there was a 60% increase in elective CS but an 85% increase in the proportion of emergency CS. Another potential driving factor behind an increased proportion of elective caesarean sections is the woman's own decision; however, this only concerns about 8% of CS in Sweden, according to a recent report published by the National Board of Health and Welfare.[209] The mother's preference could be expected to lead to an increase in elective CS, but as can be seen here, it did not explain the increase in CS deliveries overall.

But how might these changes have affected our associations? Well, for one thing, CS in itself has been found to be associated with increased risk of asthma in some studies[244-248] – but it has also been shown that this is primarily true for the emergency CS, perhaps as a consequence of stressors in the foetal environment.[208] If that is the case, the increasing proportion of emergency CS within the twin group in the time span covered by the younger cohort could have contributed to the initially higher risk of asthma in twins vs. singletons in the younger age group vs. in the older (where the crude effect was not significant). It could also explain why adjusting for this factor seemed to have different effect sizes in the different age groups.

We adjusted for mode of delivery in some of the adjusted models. For example, see model 3 and the fully adjusted model, Table 3 (older cohort) and Table 4 (younger cohort) in the *Study I* publication.[238] In summary, model 3 included mode of delivery, Apgar score, parental education, and presence of older siblings. The fully adjusted model included all of the above, but also gestational age and birth weight. Not only are gestational age and birth weight closely related, but each of these are also associated specifically with mode of delivery, because intrauterine growth restriction is a common reason to deliver a child by elective or emergency CS, and this also leads to a shift towards lower gestational age.

And let us discuss the issues of **gestational age and birth weight** further. Was the distribution of these important covariates different between the older and the younger cohort – perhaps as a consequences of changed practices in deliveries? *Figure 12*, below, shows the distribution of gestational age in twins in the older versus the younger cohort.

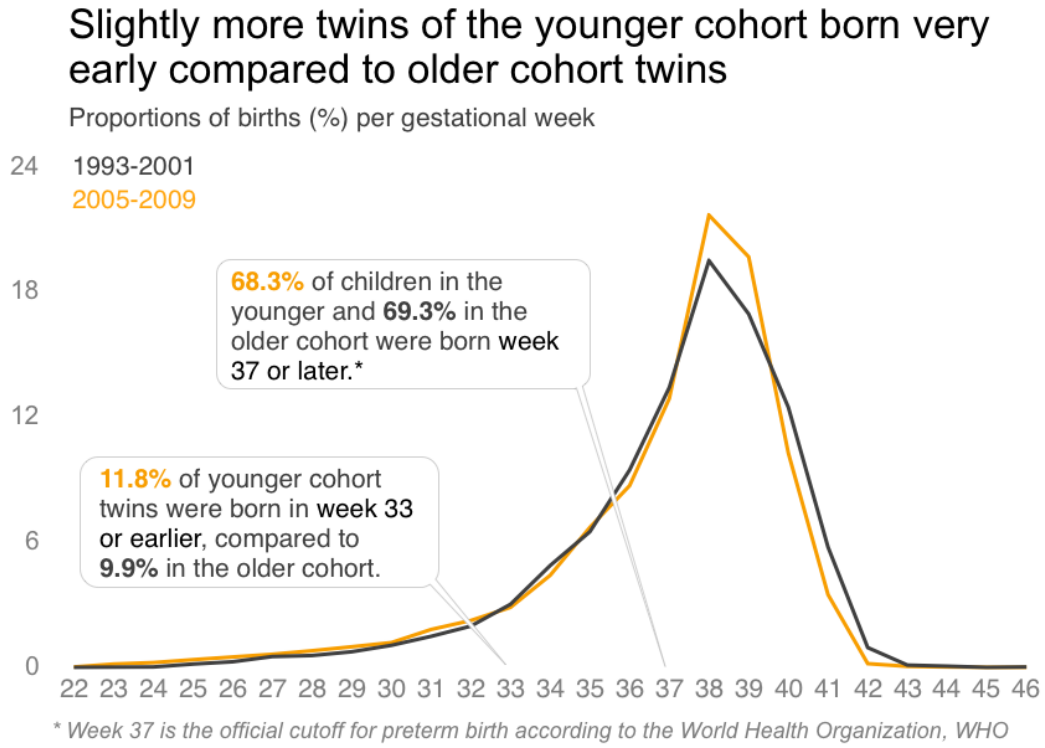

Figure 12 *Proportion of births (%, on the y axis) by gestational week in twins of the younger vs. older cohort of Study I*

There appears to be a slight shift of the orange (younger cohort) curve towards earlier gestational ages - at least in the lower and higher ranges. But, as can be seen, the absolute differences were very small; it wasn't a dramatic shift. A two-sided t-test (with unequal variance) of these groups versus one another revealed a significant difference of the mean gestational age between the younger and older cohort group (37.3 weeks in the older cohort vs. 36.9 weeks in the younger, corresponding to an absolute difference of about three days). It is however unlikely that the impact of this small difference would be large enough to make a significant contribution to the differences in the crude associations between twinship and asthma in the younger vs. older cohort.

An additional factor which has changed over time is the **structure of registers**. Before 2001, diagnoses from outpatient specialist clinics were not reported to the NPR. This means that the outpatient diagnoses of the full younger cohort were covered from birth – but the oldest participants in the older cohort did not have their outpatient diagnoses registered in the NPR until from 8 years of age.

Thus, there have been some changes over time in terms of mode of delivery, gestational age, birth weight, and methods of recording asthma diagnoses. All of these could have contributed to the different patterns of association of twinship and asthma between age groups. The use of

population-based registers also meant that information on many environmental factors – which may also have varied over time – was not available, although these could also explain some of the difference. Moreover, the de facto age difference can in itself still be an explanation. In the event of a future iteration of this study, individuals born over a longer time period and with improved register coverage could be included – this would probably reduce the influence of most of the variations discussed here.

7.1.2 Is there any support for the 'healthy twin' effect?

One potential explanation for the decreased risk of asthma in twins after adjusting for gestational age and birth weight is the so-called **healthy twin effect**.[249] The principle behind this is that if twins are more likely to be under physiological duress during pregnancy than singletons, they may thus require greater physical fitness merely to cope with the demands in utero. Adjusting for birth weight (or gestational age[250]) could make the healthy twin effect even more visible, and introduce a **birth weight paradox**.[251] This implies that adjusting for birth weight accounts for one of the primary 'unhealthy' mediating pathways between the exposure (in this case twinship) and the outcome – and that the remaining effect can to some degree be attributed to biological selection (i.e. survival despite challenges in utero).

This would suggest that twins are more fit for early delivery, owing to some special quality of twin pregnancies,[252] and perhaps explained by a physical need to be delivered earlier due to lack of space in utero. If that is the case, an interesting alternative would be to stratify twins and singletons not by their gestational age or size at birth, but by their health status. Outcomes chosen from the neonatal period have so far been the closest approach to this. Studies have compared twins and singletons matched for growth restriction[253,254] or gestational age,[255-257] but these matched studies did not find any significant differences in neonatal outcomes, including mortality, intraventricular haemorrhage, pneumothorax and early sepsis – which would have been expected in the presence of a healthy twin effect or paradox. Twins did seem to be protected to some degree from consequences of maternal pregnancy complications (gestational hypertension[258] and diabetes[259]), which could indicate something particular about twin pregnancies as such. A recent meta-analysis indicated that short time outcomes, specifically stillbirth and neonatal mortality, differ by chorionicity in twin pregnancies[260] – if possible, future studies comparing twins and singletons should take chorionicity into account.

It may be the case that twins or twin pregnancies are adapted to combat some challenges early in life, but it remains largely unclear what happens to those health adaptations over time. One Swedish study of twins born from 1932-1958 did not reveal any differences in terms of the risks of cancer, cardiovascular disease or all-cause-mortality in adult twins vs. singletons.[261] These twins and singletons were not matched for birth weight or gestational age, although low birth weight is known to increase the risk of cardiovascular disease in singleton children.[57] The apparent lack of long-term effects might indicate that differences arising from twins' and singletons' different physiological adaptations to the intrauterine environment lose relevance over time. Whether the effect is similar in terms of other

childhood (or indeed, adult) outcomes than asthma, however, remains to be assessed. Such analyses could be relevant for researchers studying other outcomes in twins.

7.2 EARLY GROWTH – SOMETHING DIFFERENT IN TWINS? *STUDY II*

In *Study II* (included in manuscript form), we modelled SITAR parameters ***a*** (size), ***b*** (tempo) and ***c*** (velocity) against asthma during childhood and adolescence. We were not able to identify any significant effects of early growth beyond a slight delay (both in terms of weight, Table 2, and height, Table 3 in the manuscript) in twins with asthma. After adjusting for factors related to foetal growth, this effect was no longer significant.

7.2.1 Closing a door and opening a window – aiming for the right time period in terms of exposure and outcome

We chose to model growth between 0-3 years of age for two reasons. By starting shortly after birth, we capture one of the earliest and most rapid phases of human growth. It is also within segments of this time window that previous studies have observed associations between early growth and asthma. *Figure 13* shows the time windows within which previous studies[95,98-110] have found associations with rapid or increased growth and asthma or wheeze.

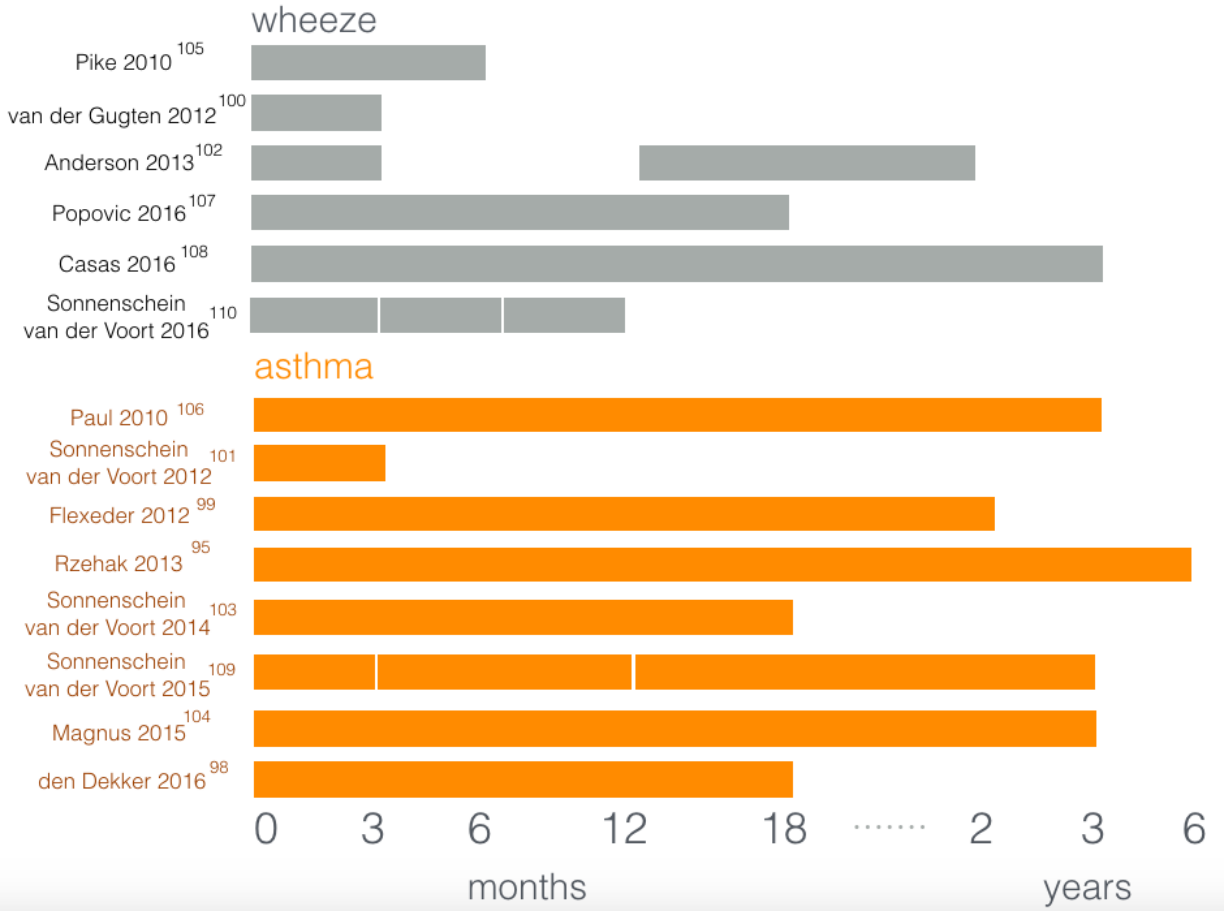

Figure 13 *Time windows (months to years of age) covered by previous studies finding associations between early childhood growth and wheeze or asthma.*

Notably, many of these studies – particularly those on wheeze[100,102,105,110] – saw their main findings between 0-3 or 0-6 months. It is possible that selecting the relatively wide age span 0-3 years for *Study II* caused us to miss some true differences that may have been present early on, but evened out over the course of the follow-up. It may also be of importance to

consider the coverage of growth data in relation to the time point(s) at which the asthma outcomes are defined. In a previous study that included asthma at several different ages (8, 14, and 17 years[109]) as outcomes, rapid weight gain during three different time intervals (0-3 months, 3-12 months, and 1-3 years) before 3 years of age were positively associated with current asthma at 8 years of age, but not at 14 years. This could imply either that early growth is less important for asthma that occurs later in childhood, or that it is the time that has passed since the rapid growth occurred in itself that makes a difference regardless of when the rapid growth took place. In *Study II*, we defined asthma as ever-asthma (occurring by 15 years of age – sometimes up to 18 years depending on the NPR and SPDR coverage) or asthma after 3 years of age, both of which will have combined early and later childhood asthma as one phenotype. Because the wheeze phenotype in CATSS only includes the children with wheeze who do not have asthma, we did not include this phenotype in *Study II*. This may have complicated comparisons to previous studies in which the asthma outcome included wheeze,[99,101,103] if it was the latter aspect of this phenotype that drove the original association.

7.2.2 Particulars of growth in twins – possible biological and behavioural origins

Another plausible explanation for our failure to find associations similar to those in previous studies would be that there is something particular about **the way twins grow**.

Let us consider the possibility that childhood growth – after birth – is recalibrated to start acting as postnatal growth regardless of which gestational week the child was born in. Specific growth charts have been constructed for the purpose of comparing the postnatal growth of prematurely born children to the interrupted foetal growth.[262] But our data did not take the twins' gestational age into account; one possibility would have been to offset the age recorded in the model by the number of weeks the birth was premature. On the other hand, if twins 'adapt' at birth to postnatal growth regardless of which gestational week they were born in (or indeed, if twin pregnancies are somehow biologically calibrated to be different and end sooner – as was discussed in connection with *Study I*), then such an approach would have been erroneous.

One potentially important factor on which we lacked information within the CATSS population was **breastfeeding**. Infants who are breastfed have slightly different early weight gain compared to formula-fed infants. Breastfed babies usually lose slightly more weight immediately after birth, and show a slightly less pronounced weight increase over the first months.[263]

It is possible that twins are breastfed less than singleton children.[264] This could in turn mean that fewer children in our study population have been breastfed than if we had conducted the same study in a singleton population. On the other hand, the prevalence of breastfeeding varies between countries,[265] and it's difficult to say what impact this would have had on the possibility of comparing our results and those of other studies.

7.2.3 Does birth weight discordance preclude co-twin analysis in terms of growth?

It is common for twins to display some degree of within-pair discordance in birth weight.[75] Potential reasons for this include differences in placentation, chorionicity and the competition for nutrition and space in utero.[182] If these factors differ between MZ and DZ pairs – and they do, because monochorionicity is more common in MZ twins[266] – the equal environments assumption may be threatened – albeit in an unconventional way. Concerns regarding the validity of this assumption usually involve excess environmental similarity between MZ as compared to DZ twins. In the case of monochorionicity and potential uneven distribution of nutrients within the twin pair during pregnancy, the opposite might be true. Selecting twin pairs discordant for birth weight would then result in a subgroup in which these differences are enhanced.

A difference in the frequency of occurrence of a phenomenon on its own, however, does not automatically translate to its biological importance. The subject also gives rise to the question of when, exactly, we should consider that differences in environment by zygosity are a potential concern. Assuming for a moment that the effects of chorionicity and placentation are strong enough to constitute a significant environmental factor with long-term consequences, it could be of importance also for studies of other exposures – not just those occurring closely to the prenatal period. Based on available knowledge of the biological impact of differences in chorionicity – the implications of which are incompletely known, especially in comparison to the many of environmental factors that twins of all zygosities share[187] – it is not possible to give an exact estimate of the impact of this potential problem on any outcome.

In summary, the assumption of equal environments within twin pregnancies of different zygosity may be a concern, but it is far from clear how great a concern it is, and this limitation is not necessarily greater in studies of growth than of other outcomes that may in some way be associated with the intrauterine environment.

7.2.4 Limitations of using registers as a data source for birth weight in terms of twins

During the 1990s, when the participants of *Study II* were born, children were not assigned their PIN until a few days after delivery. Medical Birth Register information was not formally assigned to each child until the PIN had been established. For singleton children, this would have presented no problem. But among twins, it could lead to mix-ups of birth weight within the twin pairs.

We were able to deal with this problem to some extent in *Study II* by comparing parent-reported birth weights with those recorded in the register. This was not possible in *Study I*, which did not use CATSS data. The method used to correct for this primarily relied on agreement between the parental and register-based reports, in which case the register-based value was used. An alternative might have been to rely solely on the parents –

but the potential consequences of this option have not been thoroughly investigated.

7.3 GENETICS VS. EPIGENETICS – *STUDY III* AND *IV*

Study III and *Study IV* both probe the balance between genes and environment. *Study III* primarily displays the relative contrast between the two, whereas *Study IV* illustrates a phenomenon that has long been thought to constitute a bridge between them.

In *Study III,* [267] twin models showed that additive genetics (heritability) was consistently the most important factor behind the variability of all asthma and related allergic phenotypes (*Study III*, Table 2). This isn't to say that the shared- and non-shared environment were of *no* importance. They were quite influential, too, but twin models do not allow us to identify which specific environmental factors were at work.

Although the same is true for the genetic component from a twin model, we also measured 16 selected SNPs. Of these we were able to replicate associations for six: rs3771180, rs2305480, and rs11078927 – where the minor alleles were protective – and rs12936231, rs7216389, and rs3894194 – where the minor alleles were associated with an increased risk of asthma. Altogether, these contributed a fraction of a percent of the estimated heritability in the full sample. Thus, while the individual effects of the SNPs were statistically significant, the total impact of these genetic variants was very small. While this is to be expected based on a selection of variants that was restrictive to begin with, increased knowledge regarding which specific genetic variants replicate between populations and phenotypes can still be highly beneficial – particularly in terms of extending these findings into investigations of downstream effects or potential therapeutic implications.

On that note, it is of particular interest that many of the replicated SNPs were found within the ORMDL3/GSDMB region, as our understanding of the function of these genes has increased in recent years.[268] The field still has some way to go before this knowledge can be translated to clinical practice – something that will likely require integration with subsequent studies including gene expression and epigenetic changes.

It's sometimes said that epigenetics may be one of the explanations behind the so-called **missing heritability**, but this is not entirely true. Largely, this statement probably arises from a misunderstanding of the concept of heritability. In reality, epigenetics may be part of the heritability (in the sense that it is not independent of genetic variance[126]), but should be seen as **primarily a mediating mechanism for actions of the shared- and non-shared environment.**

So what evidence did we find for such mediation in *Study IV*? At first glance, crude analyses revealed 340 CpG loci with statistically significant associations with childhood asthma – but the absolute differences in methylation between twins with and without asthma were only a few percent for each locus. Additionally, these CpGs represented several genes, but only six of these (ACOT7, IL13, KIAA0182, MAD1L1, SLC44A4 and ZFPM1[171]) replicated results from previous studies, which, moreover, have only rarely agreed with each other before. There may be several reasons for this persistent difficulty in obtaining consistent findings, one of which was elegantly illustrated in our results by the fact that adjusting for cell count

ratios weakened all of the previously found associations to the point that they were no longer significant at the false discovery rate-adjusted level. Cell counts in whole blood have to our knowledge only been accounted for in this manner in one previous study on DNA methylation in asthma.[269]

One reason for our failure to identify CpG loci that were stable following adjustment for genetic confounding is the limited number of asthma-discordant pairs and consequently lower power in the within-pair analyses. However, asthma-discordant MZ pairs are rare, and became rarer still between recruitment from CATSS and the clinical examinations in STOPPA. When co-twins who had previously been considered healthy controls were redefined as cases, previously asthma-discordant pairs became asthma-concordant.

An interesting feature of comparisons of the findings from *Study III* and *Study IV* was the lack of overlap between genetic regions of the significantly associated SNPs and CpGs. Although the SNPs in *Study III* were a highly selected group, the fact that they were hand-picked based on previous associations makes them more interesting. However, although some of the CpG sites that ranked highest in the crude model were also located on Chromosome 17 (*Study IV*, Table 1 – specifically cg05613273, cg14611258, and cg13054523), these were not located anywhere near the ORMDL3/GSDMB region. This may imply that any interaction between genetic and epigenetic variation that is relevant to this phenotype occurs between different genomic regions. Studies including both GWAS and EWAS data from the same individuals may be able to identify these interactions in the future.

7.3.1 If not epigenetics, then what?

There has to be some kind of biological difference between the twins within disease-discordant pairs - after all, if there wasn't, one twin would not be sick while the other is healthy. And yet, after correcting for genetic and cell-specific effects, we found little on the epigenetic level.

Potential explanations include that the differences are instead found in terms of gene expression, protein synthesis, metabolites, or other biomarkers.[270] All of these explanations involve biological mechanisms occurring 'one step down' from the genetic and epigenetic level. In that sense, these biomarkers are also excellent choices for validation studies of epigenetic findings. Until we have consistent, strong, replicated findings on the epigenetic level, expression and other downstream studies may have to take up the hunt on their own.

That said, there are other methods of measuring epigenetic differences – including histone acetylation[271] – and it may be that the differences that are most relevant to asthma are on this level, or that they are not covered by currently available arrays.

Finally, gene-environment interactions[272,273] and epigenetic cross-talk [274] may also be relevant in this context.

8. OVERARCHING METHODOLOGICAL CONSIDERATIONS

It just as important to be aware of the limitations of scientific work as it is to know its strengths, if not more so – for it is only by acknowledging the weaknesses in our work that we can improve upon them.

8.1 SELECTION BIAS, CONFOUNDING AND THE REMAINING MERIT OF OBSERVATIONAL STUDIES

The population recruited for *Study II* was selected via the child growth data collection, which led to an unexpected under-representation of boys.[199] Given that recruitment to all of the CATSS cohorts has involved some degree of non-response,[196] it is likely that additional factors have been unequally distributed between responders and non-responders. Although this has not been investigated in detail, the mere likelihood of unequal distribution does not automatically imply that there is selection bias, as bias requires these additional factors to be associated with both the exposure and outcome of the new study.[14]

Observational studies share an inherent weakness: we can never be completely sure whether or not an observed association might have arisen through uncontrolled, unmeasured, or residual confounding. With the help of DAGs[216] and access to data on many of the important covariates of interest, we have tried to account for as much confounding as possible within these studies. Still, some potential sources of confounding remain. For example, birth weight is only a proxy of prenatal growth and residual confounding may remain even after adjustment (*Study I* and *II*). Further, there were no data on breastfeeding, making this an uncontrolled potential mediator between infant growth and childhood asthma (*Study II*). Finally, adjusting for proportions of white blood cell populations may have obscured true differences occurring primarily within these cell types that were still disease specific (*Study IV*).

Although randomised intervention studies are designed to circumvent most of the aforementioned issues, this study design would not have been an option for the exposures studied within this thesis. Neither twinship (*Study I*), early childhood growth (*Study II*), genetic variants (*Study III*), nor DNA methylation (*Study IV*) can be randomised.

8.2 MULTIPLE TESTING, FALSE POSITIVES AND HOW THESE MAY EXPLAIN FAILED REPLICATION EFFORTS

In any study where a large number of statistical tests are run in parallel, it is possible for false positives to arise purely by chance – these are then referred to as type I errors.[275] It is the researchers who decide which statistical significance level to employ, and thereby how often we accept that this might happen. An option for dealing with this is to correct the estimates for the number of tests performed, such as by Bonferroni correction (*Study III*) or utilisation of false discovery rate-adjusted p-values (*Study IV*).

Even so, it's still possible for false positives to slip through the net. Particularly in a study based on array data (such as *Study IV*), we cannot exclude that some type I errors will remain even after correction. False positive findings in the studies from which we selected the genetic variants to feature in *Study III* could be a potential explanation for why we were not able to replicate associations for all the selected SNPs in our cohort.

8.3 MISCLASSIFICATION AND LACK OF SPECIFICITY OF THE OUTCOME

As was introduced previously, asthma is likely not a single disease – not even in childhood. But throughout this thesis, it has largely been treated as one. Retrospectively asking about 'asthma ever', which was a variable that went into construction of the outcomes in both *Study II* and *Study III*, will inevitably lead to a muddling together of all of these potentially different phenotypes. On the other hand, in *Study III*, we used a set of different definitions of asthma (including by diagnosis, medication, or questionnaire), and results were similar between these definitions.

Another inherent limitation in all of our asthma definitions is that we did not take any information about triggers for the asthma attacks into account. It is possible that asthma that is primarily allergic in nature has different risk factors and genetic components that are relevant.[276]

Still, some of the definitions are probably better than others. The medication and diagnosis outcomes used in *Study I* and *Study III* were validated against clinical records, but the self-reported asthma in CATSS-9/12, CATSS-15 / CATSS-15/DOGSS, and STOPPA has not yet been validated, although the questionnaire design is based on the International Study of Asthma and Allergies in Childhood (ISAAC) questionnaires,[277] which were.

Additionally, data collections in *Study II* and *Study IV* contained questionnaire responses from both parents and twins. In *Study IV*, we only used the parent-report of current asthma, but for *Study II* we used any report of asthma at any time up to 15 years of age, regardless of whether it had been reported by the twin or the parent. This was done because not all parents of twins at age 15 had responded, and using parent data exclusively would have meant missing data unnecessarily. Exclusively using twin data, on the other hand, might have caused us to miss out on diagnoses the twins did not themselves remember having received. Including *any* reported asthma, even if the parents and twin had all responded and disagreed, means we may have introduced some misclassification.

As diagnoses in primary care are not covered in the NPR, it may be argued that the patient population featured in the national patient register (first the inpatient register, and later also the outpatient register) is enriched to contain the more severe cases. This is probably true, but also including the medication outcomes from SPDR – which may have been prescribed anywhere – should circumvent the problem.[207]

Some studies argue that severe asthma constitutes its own phenotype.[278] In that case, results from our studies using patient register-based outcomes may be less generalisable to the

population of children with asthma as a whole. However, as the subgroup of children with very severe asthma is very small, any particular characteristics of this group should have had only minor effects on the overall results.[279]

Finally, one of the studies within this thesis (*Study III*) also examines other phenotypes besides asthma (hay fever, atopic eczema, and food allergy). All of these definitions were purely questionnaire-based and from parental report, which is a limitation. An alternative would have been to use medication proxies also for these outcomes – but an attempt to identify similar outcomes for eczema as for asthma[207] was not as successful. This may be partially attributed to the less specific prescription patterns for these diseases (especially for atopic eczema), but also to the fact that many of the medications used for allergic diseases in Sweden are sold over the counter (i.e. without the need for a prescription), and thus not captured in the SPDR.

8.4 MISCLASSIFICATION OF EXPOSURE

8.4.1 Measuring the right thing – in the right twin

During the clinical examinations in STOPPA (*Study IV*), the identity of each twin within the pair was confirmed at every step of the examination when data was recorded, and samples and tubes were clearly labelled to avoid mix-ups within the twin pair to the greatest degree possible. Since we did not have the same strict oversight over collection of the first-hand growth data measurements in *Study II* – and the twins were often measured at the same age, presumably at the same visit – here we have to rely on the accuracy of the original records. For *Study III*, it was the twins themselves who collected the saliva samples and mailed them to the KI Biobank, where they were stored until the genetic analysis could be undertaken. Although mistakes may have occurred at any stage during any of these collection processes, most likely the human error was rare and the control mechanisms in place (labels and identity checks) will have sufficed to reduce this risk to its smallest potential. In the cases where any mix-ups remain, they should be of a random (i.e. non-systematic) nature, and hopefully rare enough that their effect would be small.

8.4.2 Biological targets – testing the right tissue

Study III and *Study IV* both include analyses of biological materials: saliva (for DNA, in *Study III*) and whole blood samples (for DNA methylation, in *Study IV*). The intention in this case is to acquire a sample that is representative of the individual and, potentially, the target tissue of the disease.

For **DNA**, the process is made simpler by the fact that all nucleus-bearing cells in an individual's body share the same genetic code (in principle, at least – although over time mutations both with and without consequences may occur in individual cells). Therefore, epithelial buccal cells (such as those that can be found in a saliva sample) contain the same DNA as a cell from the blood of the same individual. That isn't to say that a saliva sample cannot also contain DNA from other organisms – indeed, it is largely for this reason that

DNA extraction kits are specific to human or, for instance, bacterial DNA.[280] It is also possible for DNA samples to be contaminated by other human DNA during the collection or handling process – although great care is taken to avoid this in laboratory settings.

With **DNA methylation**, on the other hand, the cell type matters.[127] This was why we chose to measure white blood cell counts as well as DNA methylation. Another way of handling potential variations in cell type composition is to only analyse a selected target tissue – a strategy often applied in studies of DNA methylation in asthma.[171,172,177,281,282] In many ways, this makes sense – but choosing a specific tissue type comes with its own assumptions. The researcher needs to be fairly sure that the selected tissue or cell type has clear implications in the disease under study, or else run the risk of performing the time-consuming (not to mention still rather expensive) analysis only to wonder if the actual epicentre is somewhere else.

Whole blood has commonly been analysed in DNA methylation studies, and therein lies a clear advantage to using blood samples: because they are easily accessible (especially compared to tissue biopsies), they have been used to study many different outcomes. This allows for more straightforward comparisons of results between studies, as well as for collaborations. In the long term, methylation patterns detectable in whole blood may also be of interest in clinical settings.

8.4.3 Twin pair similarity – strength or weakness?

Without the assumption that twin pairs share a large fraction of their genes and environment, twin modelling and co-twin control analyses would not be possible.

But some might argue that the fact that a twin is part of a pair might threaten the validity of reporting of outcomes and exposures within this pair. Parents of twins could be thought to over- or understate the similarity within the pairs – and the twins themselves could, too. Technically, this would constitute a form of sibling imitation or contrast effect – a phenomenon that has been described in twin studies of attention-deficit hyperactivity disorder.[283] If similarity within twin pairs *is* exaggerated, this could lead to a perceived high prevalence of the outcome in a twin population as compared to a roughly equivalent singleton population. On the other hand, if there is a counter-acting tendency to also overstate similarities in terms of *not* having the outcome, these two sides of the same mechanism may balance each other out.

We did observe relatively high prevalences of any asthma in *Study II* and *Study III*, however. Then again, these were cumulative outcomes – and prevalences from studies using similar reporting methods in Swedish singletons are also high.[24] The prevalences of our asthma outcomes in the register-based twin studies (*Study I* and *Study III*) were not so high as to make us suspect that sibling imitation or contrast effects would be a concern within health-care contexts.

Still, if sibling imitation or contrast effects occur it would lead to misclassification of either exposure or outcome. In terms of the heritability analyses in *Study III*, exaggerating similarities within the twin pairs could lead to generally higher within-pair correlations and thus an underestimate of the importance of the unique environment – the opposite would be the case if the bias were reversed (i.e. by overstating differences within the pair). In *Study II* and *Study IV*, harmonising the outcomes within the twin pair could have shifted the associations towards the null.

The discussion above is largely centred around sibling imitation or contrast effects as they would be associated with potential misclassification of the outcome. Fortunately, most of the exposures under study within this thesis were not parent-reported but rather extracted from registers (*Study I*), child health records. (*Study II*) or biological samples, (*Study IV*) and these sources should not be affected by this mechanism.

8.4.3.1 Is discordance despite a heavy load of reasons for concordance an indication of a different underlying biological mechanism?

Even in the complete absence of reporting bias in either direction, specifically selecting the pairs that are discordant in terms of exposure and outcome may be an issue in terms of generalisability. After all, there are so many reasons for twins to be similar. Yet, these particular pairs are not – at least not in terms of the outcome under study. Whatever the reason for this discordance, we cannot know for sure that the factors setting these twins apart are the same as those explaining variation in the population as a whole.

On the other hand, if the mechanism is not completely different – but instead a super-charged version of what occurs in the general population – discordant twin pairs could be seen as an ideal setting. Sometimes, representing the full source population is less important than finding the most illustrative example of a phenomenon of interest.[15]

9. CONCLUSIONS AND INTERPRETATION

The true impact of one's scientific work often cannot be assessed immediately after its conclusion, but should be viewed in the context of the on-going development within the field. In this time of specialised methodology and highly focused questions, each new finding brings its own, if small, piece to the puzzle. In the work described in this thesis, we focused on transmitted (additive genetic factors and single-nucleotide polymorphisms, *Study III*) and acquired (early childhood growth parameters, *Study II*, and DNA methylation, *Study IV*) factors in twins (in themselves the focus of *Study I*).

In the register-based study of twinship and validated asthma outcomes (***Study I***), we found that twins were at increased risk of asthma compared to singletons in early childhood, but not from 5 to 18 years of age. This could imply a true difference by age, possibly for different asthma phenotypes, or be due to changes in environment occurring between the points in time during which the two age cohorts were examined. After adjusting for perinatal covariates including gestational age, birth weight, and mode of delivery, twins were at lower risk of asthma than singletons – a fact which could be partially attributed to the underlying reasons behind early delivery in twins vs. singletons, although this merits further investigation. For the purposes of further studies of asthma in twins, however, we found the fact that the initially increased risk could be explained by the suggested mediators reassuring.

Secondly, we modelled growth during early childhood using a novel parametric method, describing specific features of changes in children's height and weight from birth to three years of age. (***Study II***) The study featured a unique combination of twin and Swedish child health record data, allowing for detailed analyses including the potential to investigate the influences of genetic and environmental confounding. At this point in childhood (or perhaps because of the broadly defined outcome that we applied) we could not find specific differences in early growth in twins with and without asthma beyond a slight delay in the timing of maximum growth velocity during this period. While genetic and environmental confounding did not appear to have influenced the delay in growth, the association disappeared following adjustment for factors associated with foetal growth, indicating that perhaps foetal growth was the true origin of the later growth pattern.

In contrast, the influence of the transmitted factors was more convincing – both for asthma (whether it was defined based on parental report or validated register-based outcomes) and related allergic diseases. (***Study III***) Although the additive genetic effects were large and highly significant in the twin models, the associations between the selected genetic variants and this phenotype only explained a fraction of the total variation. While this was not surprising given that we had chosen a small number of variants, the fact that these high-risk variants proved to be highly specific for childhood asthma was interesting – not only in the light of high heritability of all our investigated phenotypes, but because of the previously established comorbidity within this group.

Finally, more than three hundred CpG loci were significantly associated with asthma within the STOPPA cohort. (***Study IV***) But despite this relatively large number, there was limited agreement between our results and previously published data – as well as very few prior replications within the field. A possible reason for this finding may be that there have been very few studies with the potential to correct for confounding by cell type or genetics. In the discordant twin cohort design of STOPPA, we could do both – and found that adjusting for either of these factors significantly changed our results. When the prior associations weakened beyond the point of statistical significance, we concluded that DNA methylation in asthma is highly correlated to factors varying with the disease. To get to true disease-specific epigenetic features, the next generation of epigenetic studies must move beyond confounding by cell type, genetics, and other possible biological sources of confounding.

In summary, this thesis combined a unique set of data sources with novel methodologies to investigate a few well chosen angles of the two quintessential causes of every phenotype: genes and environment. For childhood asthma and the transmitted and acquired factors investigated here, the balance still seems to favour genetics. But these angles are a few of many, and the cross-sectional field of paediatric (epi)genetic epidemiology will have much to say about childhood asthma and allergies for many years to come.

Don't adventures ever have an end? I suppose not. Someone else always has to carry on the story. — J.R.R. Tolkien, The Lord of the Rings

10. IMPLICATIONS FOR FUTURE RESEARCH

We've concluded that there is still much to be done – but where to start? Based on the findings within this thesis, the following are a few questions and possible continuations which could bring the field further.

Twins were not at higher risk for asthma after adjusting for important mediators, but it is still intriguing to consider the potential that they may be healthier compared to singletons born at the same gestational age. Looking purely at the prevalence or incidence of asthma in twins vs. singletons (as in *Study I*) might only be scratching the surface. It is possible that the features of asthma are different between twins and singletons, and comparative studies of twin and singleton cohorts selected from similar populations and studied with similar methodology could be used to gain further knowledge on this matter.

Study II concerned the influence of genetic confounding on the possible association between early growth parameters and asthma. But what about the genetic influences on the early growth in itself? While it's known that outcomes of early growth are highly heritable, to my knowledge there has been no previous investigation of the heritability of features of growth on parameter-based level before. The data collected within CATSS-15 and CATSS-15/DOGSS would be an excellent setting.

In *Study III* we presented new heritability estimates for some allergic phenotypes related to asthma – some of which had few previous estimates. But all of the estimates we presented came from univariate twin models. Some bivariate models (i.e. of the shared heritability) of allergic phenotypes have been presented previously – but it could be interesting to look at the intersection between the group of related phenotypes, for example asthma, hay fever, and eczema, and also to add food allergy – a phenotype that has rarely been included in these studies previously. It might be particularly interesting to do this comparison in parallel with LD regression based on GWAS data from the same cohort.

Study IV only scratched the surface of biological confounding sources in studies of DNA methylation for paediatric outcomes. Further exploration of the association and potential genetic confounding of DNA methylation patterns in factors closely related to childhood asthma could show the width of what epigenetic changes really represent.

11. ACKNOWLEDGEMENTS

They say it takes a village to raise a child – something similar seems to be true for a new PhD. During the past years, I have had the great privilege to meet and receive guidance from a number of intelligent and inspirational individuals. Words cannot suffice to express the true depths of my gratitude to you all. Nevertheless, words are the tools at my disposal; thus I extend my sincerest thanks to:

My main supervisor, **Catarina Almqvist Malmros**. This has been a wonderful journey, and you are an extraordinary guide. While writing these acknowledgements, I've attempted to pinpoint all the things you've done so that I could describe them in detail. But the trademark of a great leader is the capability to meet people exactly where they are – and then help them find the tools to move forward. I've changed over the course of the years we've worked together – as has your approach to teaching me. In the end I think the beauty of it all is that you've always given me exactly what I've needed. Deciding to work with you is undoubtedly the best choice I have made in academia – I would recommend any other student to do the same in a heartbeat.

My co-supervisors:

Paul Lichtenstein, from discussions with whom I have always come away with valuable insights. Paul, you have this way of really listening and identifying the core of any problem – an instant acuity that gives an edge to any project you're involved in. Thank you also for heading the department and ensuring that MEB is certainly one of the best places possible to pursue a PhD.

Patrik Magnusson: Your enthusiasm for research in general and genetics in particular is contagious. Together with a knack for explaining the most complex of concepts in ways that make them seem easy, the sum of your presence has encouraged me to continuous improvement.

Göran Pershagen, who has brought wise insights and clinical as well as extensive epidemiological experience. Even though your office is across the road and in another department, your advice has never been far away when I've needed it.

When I first came to MEB, I was a near total beginner when it came to statistical analysis. **Cecilia Lundholm** saved my senses on more than one occasion when the analysis software Sas or Stata seemed to be against me. With you, Cecilia, no question – no matter how basic – ever seemed silly. Thanks to your guidance, I survived survival analysis and managed to do more than play with the SITAR model. For bioinformatic support and guidance in the strange new world of R, **Robert Karlsson** was a blessing. If not for you, Robert, I would probably still be stuck battling error messages and writing ridiculously long-winded code to solve

surprisingly simple issues. I also had invaluable methodological support from **Ralf Kuja-Halkola**, who taught me everything I know about twin models.

My other **co-authors** on papers within and outside this thesis: **Andreas Kruse, Anna Zettergren, Anne Örtqvist, Brett Toelle, Brian D'Onofrio, Chris Cowell, Clara Hellner Gumpert, David Andersson, Emma Frans, Erik Melén, Frances Garden, Frida Svensdotter, Guy Marks, Henrik Anckarsäter, Henrik Olsson, Maria Movin,** and **Sebastian Lundström.** Thank you for sharing your knowledge with me as well as each other. Collaboration is the key to success.

My external mentor **Hans Berglund**, who challenged me to consider my clinical role in relation to my research, and to reflect on the many facets of the combined role of an MD / PhD.

Past and present members of the Almqvist Malmros research group, including **Anne Örtqvist, Tong Gong, Gustaf Rejnö, Cecilia Lundholm, Elin Dahlén, Jennifer Protudjer, Kirsten Holmberg, Björn Nordlund, Anna Hedman, Bronwyn Haasdyk Brew, Sandra Ganrud Tedner**, and all our **degree project students**. Thank you for the stimulating meetings and rewarding work environment!

Behind every successful study is a number of invaluable contributors who do not get their names on the author list. The **twins and parents participating in CATSS** and **STOPPA** have generously given their time and their samples. The large amounts of new data generated from these studies were expertly handled by the DBA, IT department, and data collectors at MEB. **Camilla Palm**, **Isabelle Kizling,** and **Rozita Broumandi** were my magical database gurus. **Rikard Öberg** showed me the ropes with the skjold server. **Gerd Agerberg** was essential for the growth data entry (*Study II*) and the STOPPA questionnaires (*Study IV*) – it would have taken ages without your help! **Caroline Almqvist, Jakob Malmros, Hanna Malmros**, **Patrik Holmberg, Tong Gong,** and **Anne Örtqvist** also made valuable contributions to the data entry in *Study II*. The **staff at county council archives** all over Sweden sent or kindly allowed me to come collect copies of the child health records.

Additionally, for their work during the data collection phase of STOPPA I would like to thank **Anne Örtqvist**, **Karin Dellenvall, Malena Kjellén, Jeanette Tedborn, Monika Nordenbrand, Faiza Abdillahi, Eva Thunberg, Fanny Ståhl**, and all the nurses and collaborators at the clinics and test centres in Stockholm, Gothenburg, Karlstad, Linköping, Lund, Lomma, Växjö, and Umeå.

The laboratory analysis and initial quality control of the DNA methylation analyses in *Study IV* were excellently performed by **Lovisa Reinius, Gunnar Falk, Astrid Fungmark,** and **Kristina Duvefelt** at the Mutation Analysis Facility (MAF). **Camilla Lagerberg** and **Tove Rylander Rudqvist** at the KI Biobank administered the DNA extraction.

To some of the postdocs who have graced MEB: **Jennifer Protudjer**; just before my admission seminar, you told me to enjoy the ride. In retrospect, this was probably the best single piece of advice anyone has given me concerning my PhD. And I certainly did! **Sarah Bergen** and **Anna Kähler**, the two of you have been somewhat of my postdoc role models – I hope you'll let me continue to pick your minds in the future.

MEB is a great place for PhD students. This is in many ways thanks to **Camilla Ahlqvist,** our excellent education administrator. We would all be lost in a sea of forms without you.

The PubMEB gang of 2015: **Elisabeth Dahlqwist, Robert Karlsson, Andreas Jangmo, Henrik Olsson, Johanna Holm, Andreas Karlsson, Favelle Lamb,** and **Emilio Ugalde**. We've had some great times – I hope you'll join me for one more party!

My past and present colleagues in the steering group of the Centre for Allergy Research (CfA) PhD Student and Postdoc Network: **Jennifer Protudjer, Jeanette Grundström, Jesper Säfholm, Jonas Binnmyr, Martijn Manson, Nicole Marquardt,** and **Olivia Larsson.**

The students I've helped teach: **David Andersson, Andreas Kruse, Erik Håkansson, Ida Manninen**, and all the VetU students from the medical program. Thank you for helping me see that teaching is something I truly enjoy.

More or less recent PhD graduates of MEB: **Adina Feldman** – thank you for shining so brightly, like a beacon of inspiration for what a PhD education at MEB could be. I hope I managed to do your hopes for the future of MEB and the PhD student group justice! **Anne Örtqvist –** always one step ahead – thank you for being a constant source of motivation! It wouldn't have been the same without you, either. **Lisa Möller** – for encouraging me to diversify my education and my career. **Sandra Eloranta –** you trigger my impulses to pursue cool and slightly crazy challenges. The world needs more like you. **Therese Ljung** – you were a great companion and sounding board at the top of the PhD group – I really enjoyed working with you. **Tong Gong** – your bright laughter and helpful spirit were pleasant additions to my days. **Tracy Peters** – for bringing your upbeat American vibes to improve the mood – whether over tea, dinner, or drinks. Thank you also to **Alexander Viktorin, Amir Sariaslan, Andrea Ganna, Anna Johansson, Emma Frans, Iffat Rahman, Karin Sundström, Katherine Kasiman, Lovisa Brehmer née Högberg, Maria Schelin née Sandberg, Martin Cederlöf, Martin Fransson, Miriam Elfström, Myeongjee Lee, Qi Chen, Ralf Kuja-Halkola, Robert Karlsson, Sara Raposo née Christensen, Stephanie Bonn, Therese Andersson, Thomas Frisell,** and **Viktoria Johansson**, who went ahead.

The current PhD students at MEB, including **Agnes Ohlsson Gotby**, **Andreas Jangmo, Andreas Karlsson, Caroline Weibull, Carolyn Cesta, Daniela Mariosa, Elisa Longinetti, Elisabeth Dahlqwist, Emilio Ugalde, Favelle Lamb, Fei Yang, Frida Lundberg,**

Hannah Bower, Ida Karlsson, Isabell Brikell, Isabella Ekheden, Johanna Holm, Johanna Sieurin, Kathleen Bokenberger, Laura Ghirardi, Linda Abrahamsson, Malin Ericsson, Nelson Ndegwa, and **Shuyang Yao**. You may feel like the day is far away, but you'll be next. I won't pretend the final countdown isn't the most hectic part of the road – but it's probably the most enjoyable one, too. Don't forget to have fun!

Among the current MEB PhD students, a special mention to **Johanna Holm** – for the best fika-walks, conversations about life, academia, and everything – and for refreshing my thoughts at some of the points when I've needed it the most. You've made me better – as a researcher, leader, and person. *Thank you.*

My friends from around the globe: **Amanda Roberts, Anett Wargmo, Ashley Dawn Lu, Chelsea Heidt, Clarissa Alberts, Hannah Stevens, Jacqueline Cooke, Jenny Hiltunen, Josephine Hellberg, Julia Svechnikova, Sophie Buchanan, Tammar Gino,** and **Ziang Her.** Distance means nothing when one has friends like you.

I am incredibly lucky to come from a large and wonderful **extended family**. You have always been there to celebrate all of my achievements – I hope you will particularly enjoy sharing this one! Somehow, my godmother **Kerstin Nordlund-Malmegård** seemed to know I would combine medical school and research long before I did. She and her twin sister **Ulla Nordlund-Gustafsson** have always encouraged me to pursue my goals.

My aunt and godmother **Viveca Ullemar Friberg** does not just collect gems – she *is* one.

My maternal **grandmother Rut**, thank you for pointing out that I work too much (you're probably right) and for instilling a rather fervent hope in the power of genetics because I can think of no better way of growing old than the way you have. Even I can hardly believe you're really 99!

My **brother Leo**: growing up alongside you has made me who I am. I am proud to be able to call myself your sister, and I admire you more than you know. You're the best. ♥

My father **Bengt**, who showed me the sky is not the limit, but something very much within reach. Thank you for indulging my incessant ambition – your support has been the wind beneath *my* wings. And my mother **Carina**, who put science in my DNA. While *my* thesis is not going to result in a second Ullemar's formula (probably just as well – after all, what would they name it?) it *is* going to bring another Dr. Ullemar into the world. And yet, none of what I've accomplished would have been possible if it had not been for the first one. Mamma, this is just as much your work as it is mine.

To all current and future bearers of the Ullemar name: you can do anything you set your mind to. Reach for the stars – I'll be cheering you on.

12. REFERENCES

1. Hsia CC, Schmitz A, Lambertz M, Perry SF, Maina JN. Evolution of air breathing: oxygen homeostasis and the transitions from water to land and sky. *Compr Physiol.* 2013;3(2):849-915.

2. Morrisey EE, Hogan BL. Preparing for the first breath: genetic and cellular mechanisms in lung development. *Dev Cell.* 2010;18(1):8-23.

3. WHO. Asthma Fact Sheet N°307. http://www.who.int/mediacentre/factsheets/fs307/en/. Accessed 2013-12-22.

4. Pedersen SE, Hurd SS, Lemanske RF, Jr., et al. Global strategy for the diagnosis and management of asthma in children 5 years and younger. *Pediatr Pulmonol.* 2011;46(1):1-17.

5. Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma P. The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy.* 2004;59(5):469-478.

6. Anandan C, Nurmatov U, van Schayck OC, Sheikh A. Is the prevalence of asthma declining? Systematic review of epidemiological studies. *Allergy.* 2010;65(2):152-167.

7. Beasley R, Semprini A, Mitchell EA. Risk factors for asthma: is prevention possible? *Lancet.* 2015;386(9998):1075-1085.

8. Dolinoy DC, Weidman JR, Jirtle RL. Epigenetic gene regulation: linking early developmental environment to adult disease. *Reprod Toxicol.* 2007;23(3):297-307.

9. Meulemans WJ, Lewis CM, Boomsma DI, et al. Genetic modelling of dizygotic twinning in pedigrees of spontaneous dizygotic twins. *Am J Med Genet.* 1996;61(3):258-263.

10. Garattini S, Jakobsen JC, Wetterslev J, et al. Evidence-based clinical practice: Overview of threats to the validity of evidence and how to minimise them. *Eur J Intern Med.* 2016;32:13-21.

11. Rothman KJ. *Modern epidemiology.* 3 ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

12. VanderWeele TJ, Hernan MA. Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs. *Am J Epidemiol.* 2012;175(12):1303-1310.

13. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol.* 1990;43(1):87-91.

14. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004;15(5):615-625.

15. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol.* 2013;42(4):1012-1014.

16. Sears MR, Greene JM, Willan AR, et al. A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. *N Engl J Med.* 2003;349(15):1414-1422.

17. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. *N Engl J Med.* 1995;332(3):133-138.

18. Morgan WJ, Stern DA, Sherrill DL, et al. Outcome of asthma and wheezing in the first 6 years of life: follow-up through adolescence. *Am J Respir Crit Care Med.* 2005;172(10):1253-1258.

19. Henderson J, Granell R, Heron J, et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax.* 2008;63(11):974-980.

20. Spycher BD, Silverman M, Brooke AM, Minder CE, Kuehni CE. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *Eur Respir J.* 2008;31(5):974-981.

21. Savenije OE, Granell R, Caudri D, et al. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *Journal of Allergy and Clinical Immunology.* 2011;127(6):1505-1512.e1514.

22. Collins SA, Pike KC, Inskip HM, et al. Validation of novel wheeze phenotypes using longitudinal airway function and atopic sensitization data in the first 6 years of life: evidence from the Southampton Women's survey. *Pediatr Pulmonol.* 2013;48(7):683-692.

23. Duijts L, Granell R, Sterne JA, Henderson AJ. Childhood wheezing phenotypes influence asthma, lung function and exhaled nitric oxide fraction in adolescence. *Eur Respir J.* 2016;47(2):510-519.

24. Hallberg J, Thunqvist P, Schultz ES, et al. Asthma phenotypes and lung function up to 16 years of age-the BAMSE cohort. *Allergy.* 2015;70(6):667-673.

25. Brand PL, Caudri D, Eber E, et al. Classification and pharmacological treatment of preschool wheezing: changes since 2008. *Eur Respir J.* 2014;43(4):1172-1177.

26. Depner M, Fuchs O, Genuneit J, et al. Clinical and epidemiologic phenotypes of childhood asthma. *Am J Respir Crit Care Med.* 2014;189(2):129-138.

27. Bateman ED, Hurd SS, Barnes PJ, et al. Global strategy for asthma management and prevention: GINA executive summary. *Eur Respir J.* 2008;31(1):143-178.

28. Just J, Saint Pierre P, Amat F, et al. What lessons can be learned about asthma phenotypes in children from cohort studies? *Pediatr Allergy Immunol.* 2015;26(4):300-305.

29. Granell R, Henderson AJ, Timpson N, et al. Examination of the relationship between variation at 17q21 and childhood wheeze phenotypes. *J Allergy Clin Immunol.* 2013;131(3):685-694.

30. Evans CM, Kim K, Tuvim MJ, Dickey BF. Mucus hypersecretion in asthma: causes and effects. *Curr Opin Pulm Med.* 2009;15(1):4-11.

31. National Asthma Education and Prevention Program TEPotDaMoA. *Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Section 2, Definition, Pathophysiology and Pathogenesis of Asthma, and Natural History of Asthma.* Bethesda (MD)2007.

32. Reddel HK, Bateman ED, Becker A, et al. A summary of the new GINA strategy: a roadmap to asthma control. *Eur Respir J.* 2015;46(3):622-639.

33. Abbas AK, Lichtman AH. Basic Immunology - Functions and Disorders of the Immune System. 3 ed. Philadelphia: Saunders; 2009:8-11.

34. Romagnani S. T-cell subsets (Th1 versus Th2). *Ann Allergy Asthma Immunol.* 2000;85(1):9-18; quiz 18, 21.

35. Kim Y, Lee S, Kim YS, et al. Regulation of Th1/Th2 cells in asthma development: a mathematical model. *Math Biosci Eng.* 2013;10(4):1095-1133.

36. Zhu M, Liang Z, Wang T, Chen R, Wang G, Ji Y. Th1/Th2/Th17 cells imbalance in patients with asthma with and without psychological symptoms. *Allergy Asthma Proc.* 2016;37(2):148-156.

37. Kumar A, Das S, Agrawal A, Mukhopadhyay I, Ghosh B. Genetic association of key Th1/Th2 pathway candidate genes, IRF2, IL6, IFNGR2, STAT4 and IL4RA, with atopic asthma in the Indian population. *J Hum Genet.* 2015;60(8):443-448.

38. Ballardini N, Kull I, Lind T, et al. Development and comorbidity of eczema, asthma and rhinitis to age 12: data from the BAMSE birth cohort. *Allergy.* 2012;67(4):537-544.

39. Garcia-Aymerich J, Benet M, Saeys Y, et al. Phenotyping asthma, rhinitis and eczema in MeDALL population-based birth cohorts: an allergic comorbidity cluster. *Allergy.* 2015;70(8):973-984.

40. Burte E, Bousquet J, Varraso R, et al. Characterization of Rhinitis According to the Asthma Status in Adults Using an Unsupervised Approach in the EGEA Study. *PLoS One.* 2015;10(8):e0136191.

41. Mirabelli MC, Hsu J, Gower WA. Comorbidities of asthma in U.S. children. *Respir Med.* 2016;116:34-40.

42. Lowe AJ, Carlin JB, Bennett CM, et al. Do boys do the atopic march while girls dawdle? *J Allergy Clin Immunol.* 2008;121(5):1190-1195.

43. Dharmage SC, Lowe AJ, Matheson MC, Burgess JA, Allen KJ, Abramson MJ. Atopic dermatitis and the atopic march revisited. *Allergy.* 2014;69(1):17-27.

44. Spergel JM. From atopic dermatitis to asthma: the atopic march. *Ann Allergy Asthma Immunol.* 2010;105(2):99-106; quiz 107-109, 117.

45. Belgrave DC, Granell R, Simpson A, et al. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med.* 2014;11(10):e1001748.

46. Alduraywish SA, Lodge CJ, Campbell B, et al. The march from early life food sensitization to allergic disease: a systematic review and meta-analyses of birth cohort studies. *Allergy.* 2016;71(1):77-89.

47. Belgrave DC, Simpson A, Semic-Jusufagic A, et al. Joint modeling of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. *J Allergy Clin Immunol.* 2013;132(3):575-583 e512.

48. de Amici M, Ciprandi G. The age impact on serum total and allergen-specific IgE. *Allergy Asthma Immunol Res.* 2013;5(3):170-174.

49. Fagnani C, Annesi-Maesano I, Brescianini S, et al. Heritability and shared genetic effects of asthma and hay fever: an Italian study of young twins. *Twin Res Hum Genet.* 2008;11(2):121-131.

50. Ortqvist AK, Lundholm C, Carlstrom E, Lichtenstein P, Cnattingius S, Almqvist C. Familial factors do not confound the association between birth weight and childhood asthma. *Pediatrics.* Vol 124. United States2009:e737-743.

51. Lundholm C, Ortqvist AK, Lichtenstein P, Cnattingius S, Almqvist C. Impaired fetal growth decreases the risk of childhood atopic eczema: a Swedish twin study. *Clin Exp Allergy.* Vol 40. England2010:1044-1053.

52. Barnish MS, Tagiyeva N, Devereux G, Aucott L, Turner S. Diverging prevalences and different risk factors for childhood asthma and eczema: a cross-sectional study. *BMJ Open.* 2015;5(6):e008446.

53. Kantor R, Kim A, Thyssen J, Silverberg JI. Association of atopic dermatitis with smoking: A systematic review and meta-analysis. *J Am Acad Dermatol.* 2016.

54. Ronmark EP, Ekerljung L, Mincheva R, et al. Different risk factor patterns for adult asthma, rhinitis and eczema: results from West Sweden Asthma Study. *Clin Transl Allergy.* 2016;6:28.

55. Simpson AB, Yousef E, Hossain J. Association between peanut allergy and asthma morbidity. *J Pediatr.* 2010;156(5):777-781, 781 e771.

56. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet.* 1986;1(8489):1077-1081.

57. Barker DJ. Fetal origins of coronary heart disease. *BMJ.* 1995;311(6998):171-174.

58. Krauss-Etschmann S, Bush A, Bellusci S, et al. Of flies, mice and men: a systematic approach to understanding the early life origins of chronic lung disease. *Thorax.* 2013;68(4):380-384.

59. Shaheen SO, Sterne JA, Montgomery SM, Azima H. Birth weight, body mass index and asthma in young adults. *Thorax.* 1999;54(5):396-402.

60. Mu M, Ye S, Bai MJ, et al. Birth weight and subsequent risk of asthma: a systematic review and meta-analysis. *Heart Lung Circ.* 2014;23(6):511-519.

61. Liu X, Olsen J, Agerbo E, et al. Birth weight, gestational age, fetal growth and childhood asthma hospitalization. *Allergy Asthma Clin Immunol.* 2014;10(1):13.

62. Kindlund K, Thomsen SF, Stensballe LG, et al. Birth weight and risk of asthma in 3-9-year-old twins: exploring the fetal origins hypothesis. *Thorax.* Vol 65. England2010:146-149.

63. Tedner SG, Ortqvist AK, Almqvist C. Fetal growth and risk of childhood asthma and allergic disease. *Clin Exp Allergy.* 2012;42(10):1430-1447.

64. Davidson R, Roberts SE, Wotton CJ, Goldacre MJ. Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: evidence from the Oxford record linkage study. *Bmc Pulmonary Medicine.* 2010;10.

65. Dik N, Tate RB, Manfreda J, Anthonisen NR. Risk of physician-diagnosed asthma in the first 6 years of life. *Chest.* 2004;126(4):1147-1153.

66. Kallen B, Finnstrom O, Nygren KG, Otterblad Olausson P. Association between preterm birth and intrauterine growth retardation and child asthma. *Eur Respir J.* 2013;41(3):671-676.

67. Neuman A, Hohmann C, Orsini N, et al. Maternal smoking in pregnancy and asthma in preschool children: a pooled analysis of eight birth cohorts. *Am J Respir Crit Care Med.* 2012;186(10):1037-1043.

68. McKeever TM, Lewis SA, Smith C, et al. Siblings, multiple births, and the incidence of allergic disease: a birth cohort study using the West Midlands general practice research database. *Thorax.* 2001;56(10):758-762.

69. Almqvist C, Olsson H, Fall T, Lundholm C. Sibship and risk of asthma in a total population: A disease comparative approach. *J Allergy Clin Immunol.* 2016.

70. Fall T, Lundholm C, Ortqvist AK, et al. Early Exposure to Dogs and Farm Animals and the Risk of Childhood Asthma. *JAMA Pediatr.* 2015;169(11):e153219.

71. Thacher JD, Gruzieva O, Pershagen G, et al. Pre- and postnatal exposure to parental smoking and allergic disease through adolescence. *Pediatrics.* 2014;134(3):428-434.

72. Gern JE, Rosenthal LA, Sorkness RL, Lemanske RF, Jr. Effects of viral respiratory infections on lung development and childhood asthma. *J Allergy Clin Immunol.* 2005;115(4):668-674; quiz 675.

73. Ortqvist AK, Lundholm C, Kieler H, et al. Antibiotics in fetal and early life and subsequent childhood asthma: nationwide population based study with sibling analysis. *BMJ.* 2014;349:g6979.

74. Haram K, Softeland E, Bukowski R. Intrauterine growth restriction. *Int J Gynaecol Obstet.* 2006;93(1):5-12.

75. Puccio G, Giuffre M, Piccione M, Piro E, Malerba V, Corsello G. Intrauterine growth pattern and birthweight discordance in twin pregnancies: a retrospective study. *Ital J Pediatr.* 2014;40:43.

76. Nguyen HT, Eriksson B, Nguyen LT, et al. Physical growth during the first year of life. A longitudinal study in rural and urban areas of Hanoi, Vietnam. *BMC Pediatr.* 2012;12:26.

77. Williams JP. Catch-up growth. *J Embryol Exp Morphol.* 1981;65 Suppl:89-101.

78. Ong KK, Ahmed ML, Emmett PM, Preece MA, Dunger DB. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. *BMJ.* 2000;320(7240):967-971.

79. Mihrshahi S, Battistutta D, Magarey A, Daniels LA. Determinants of rapid weight gain during infancy: baseline results from the NOURISH randomised controlled trial. *BMC Pediatr.* 2011;11:99.

80. Figueroa-Munoz JI, Chinn S, Rona RJ. Association between obesity and asthma in 4-11 year old children in the UK. *Thorax.* 2001;56(2):133-137.

81. Chinn S, Rona RJ. Can the increase in body mass index explain the rising trend in asthma in children? *Thorax.* 2001;56(11):845-850.

82. Castro-Rodriguez JA, Holberg CJ, Morgan WJ, Wright AL, Martinez FD. Increased incidence of asthmalike symptoms in girls who become overweight or obese during the school years. *Am J Respir Crit Care Med.* 2001;163(6):1344-1349.

83. Gold DR, Damokosh AI, Dockery DW, Berkey CS. Body-mass index as a predictor of incident asthma in a prospective cohort of children. *Pediatr Pulmonol.* 2003;36(6):514-521.

84. Mamun AA, Lawlor DA, Alati R, O'Callaghan MJ, Williams GM, Najman JM. Increasing body mass index from age 5 to 14 years predicts asthma among adolescents: evidence from a birth cohort study. *Int J Obes (Lond).* 2007;31(4):578-583.

85. Hong SJ, Lee MS, Lee SY, et al. High body mass index and dietary pattern are associated with childhood asthma. *Pediatr Pulmonol.* 2006;41(12):1118-1124.

86. Sithole F, Douwes J, Burstyn I, Veugelers P. Body mass index and childhood asthma: a linear association? *J Asthma.* 2008;45(6):473-477.

87. Ahmadiafshar A, Tabbekhha S, Mousavinasab N, Khoshnevis P. Relation between asthma and body mass index in 6-15 years old children. *Acta Med Iran.* 2013;51(9):615-619.

88. Granell R, Henderson AJ, Evans DM, et al. Effects of BMI, fat mass, and lean mass on asthma in childhood: a Mendelian randomization study. *PLoS Med.* 2014;11(7):e1001669.

89. Scholtens S, Wijga AH, Seidell JC, et al. Overweight and changes in weight status during childhood in relation to asthma symptoms at 8 years of age. *J Allergy Clin Immunol.* 2009.

90. Magnusson JO, Kull I, Mai XM, Wickman M, Bergstrom A. Early childhood overweight and asthma and allergic sensitization at 8 years of age. *Pediatrics.* 2012;129(1):70-76.

91. Loid P, Goksor E, Alm B, et al. A persistently high body mass index increases the risk of atopic asthma at school age. *Acta Paediatr.* 2015;104(7):707-712.

92. Wandalsen GF, Borges LV, Barroso N, et al. Gender differences in the relationship between body mass index (BMI) changes and the prevalence and severity of wheezing and asthma in the first year of life. *Allergol Immunopathol (Madr).* 2015.

93. Porter M, Wegienka G, Havstad S, et al. Relationship between childhood body mass index and young adult asthma. *Ann Allergy Asthma Immunol.* 2012;109(6):408-411 e401.

94. Ziyab AH, Karmaus W, Kurukulaaratchy RJ, Zhang H, Arshad SH. Developmental trajectories of Body Mass Index from infancy to 18 years of age: prenatal determinants and health consequences. *J Epidemiol Community Health.* 2014;68(10):934-941.

95. Rzehak P, Wijga AH, Keil T, et al. Body mass index trajectory classes and incident asthma in childhood: results from 8 European Birth Cohorts--a Global Allergy and Asthma European Network initiative. *J Allergy Clin Immunol.* 2013;131(6):1528-1536.

96. Belfort MB, Cohen RT, Rhein LM, McCormick MC. Preterm infant growth and asthma at age 8 years. *Arch Dis Child Fetal Neonatal Ed.* 2016;101(3):F230-234.

97. Kreissl S, Radon K, Dressel H, et al. Body mass index change and atopic diseases are not always associated in children and adolescents. *Ann Allergy Asthma Immunol.* 2014;113(4):440-444.e441.

98. den Dekker HT, Sonnenschein-van der Voort AM, de Jongste JC, et al. Early growth characteristics and the risk of reduced lung function and asthma: A meta-analysis of 25,000 children. *J Allergy Clin Immunol.* 2016;137(4):1026-1035.

99. Flexeder C, Thiering E, Bruske I, et al. Growth velocity during infancy and onset of asthma in school-aged children. *Allergy.* 2012;67(2):257-264.

100. van der Gugten A, Koopman M, Evelein A, Verheij T, Uiterwaal C, van der Ent C. Rapid early weight gain is associated with wheeze and reduced lung function in childhood. *Eur Respir J.* 2012;39(2):403-410.

101. Sonnenschein-van der Voort AM, Jaddoe VW, Raat H, et al. Fetal and infant growth and asthma symptoms in preschool children: the Generation R Study. *Am J Respir Crit Care Med.* 2012;185(7):731-737.

102. Anderson EL, Fraser A, Martin RM, et al. Associations of postnatal growth with asthma and atopy: the PROBIT Study. *Pediatr Allergy Immunol.* 2013;24(2):122-130.

103. Sonnenschein-van der Voort AM, Arends LR, de Jongste JC, et al. Preterm birth, infant weight gain, and childhood asthma risk: a meta-analysis of 147,000 European children. *J Allergy Clin Immunol.* 2014;133(5):1317-1329.

104. Magnus MC, Stigum H, Haberg SE, Nafstad P, London SJ, Nystad W. Peak weight and height velocity to age 36 months and asthma development: the Norwegian Mother and Child Cohort Study. *PLoS One.* 2015;10(1):e0116362.

105. Pike KC, Crozier SR, Lucas JS, et al. Patterns of fetal and infant growth are related to atopy and wheezing disorders at age 3 years. *Thorax.* 2010;65(12):1099-1106.

106. Paul IM, Camera L, Zeiger RS, et al. Relationship between infant weight gain and later asthma. *Pediatr Allergy Immunol.* Vol 21. England: 2009 John Wiley & Sons A/S; 2010:82-89.

107. Popovic M, Pizzi C, Rusconi F, et al. Infant weight trajectories and early childhood wheezing: the NINFEA birth cohort study. *Thorax.* 2016.

108. Casas M, den Dekker HT, Kruithof CJ, et al. Early childhood growth patterns and school-age respiratory resistance, fractional exhaled nitric oxide and asthma. *Pediatr Allergy Immunol.* 2016.

109. Sonnenschein-van der Voort A, Howe LD, Granell R, et al. Influence of childhood growth on asthma and lung function in adolescence. *J Allergy Clin Immunol.* 2015;135(6):1435-1443 e1437.

110. Sonnenschein-van der Voort AM, Gaillard R, de Jongste JC, Hofman A, Jaddoe VW, Duijts L. Foetal and infant growth patterns, airway resistance and school-age asthma. *Respirology.* 2016;21(4):674-682.

111. Matos SM, Jesus SR, Saldiva SR, et al. Weight gain in the first two years of life, asthma and atopy: the SCAALA cohort study. *Public Health Nutr.* 2014;17(11):2537-2545.

112. Mebrahtu TF, Feltbower RG, Parslow RC. Effects of birth weight and growth on childhood wheezing disorders: findings from the Born in Bradford Cohort. *BMJ Open.* 2015;5(11):e009553.

113. Silventoinen K, Sammalisto S, Perola M, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* 2003;6(5):399-408.

114. Silventoinen K, Bartels M, Posthuma D, et al. Genetic regulation of growth in height and weight from 3 to 12 years of age: a longitudinal study of Dutch twin children. *Twin Res Hum Genet.* 2007;10(2):354-363.

115. Tarnoki AD, Tarnoki DL, Medda E, et al. Bioimpedance analysis of body composition in an international twin cohort. *Obes Res Clin Pract.* 2014;8(3):e201-298.

116. Lichtenstein P, Svartengren M. Genes, environments, and sex: factors of importance in atopic diseases in 7-9-year-old Swedish twins. *Allergy.* 1997;52(11):1079-1086.

117. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature.* 1953;171(4356):737-738.

118. Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. *PLoS Biol.* 2007;5(10):e254.

119. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science.* 1998;280(5366):1077-1082.

120. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012;13(7):484-492.

121. Lo R, Weksberg R. Biological and biochemical modulation of DNA methylation. *Epigenomics.* 2014;6(6):593-602.

122. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002;16(1):6-21.

123. Gibbs JR, van der Brug MP, Hernandez DG, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* 2010;6(5):e1000952.

124. Zhang D, Cheng L, Badner JA, et al. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet.* 2010;86(3):411-419.

125. Bell JT, Pai AA, Pickrell JK, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* 2011;12(1):R10.

126. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife.* 2013;2:e00523.

127. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One.* 2012;7(7):e41361.

128. Daly AK, Day CP. Candidate gene case-control association studies: advantages and potential pitfalls. *Br J Clin Pharmacol.* 2001;52(5):489-499.

129. Moffatt MF, Cookson WO. Linkage and candidate gene studies in asthma. *Am J Respir Crit Care Med.* 1997;156(4 Pt 2):S110-112.

130. Rogers AJ, Raby BA, Lasky-Su JA, et al. Assessing the reproducibility of asthma candidate gene associations, using genome-wide data. *Am J Respir Crit Care Med.* 2009;179(12):1084-1090.

131. Michel S, Liang L, Depner M, et al. Unifying candidate gene and GWAS Approaches in Asthma. *PLoS One.* 2010;5(11):e13894.

132. Topol EJ, Murray SS, Frazer KA. The genomics gold rush. *JAMA.* 2007;298(2):218-221.

133. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA.* 2008;299(11):1335-1344.

134. Ingelsson E. Large-scale genome-wide association studies consortia: blessing, burden, or necessity? *Circ Cardiovasc Genet.* Vol 3. United States2010:396-398.

135. Bibikova M, Le J, Barnes B, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics.* 2009;1(1):177-200.

136. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics.* 2011;98(4):288-295.

137. Sandoval J, Heyn H, Moran S, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics.* 2011;6(6):692-702.

138. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics.* 2016;8(3):389-399.

139. Sun Z, Chai HS, Wu Y, et al. Batch effect correction for genome-wide methylation data with Illumina Infinium platform. *BMC Med Genomics.* 2011;4:84.

140. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis.* 1990;142(6 Pt 1):1351-1358.

141. Harris JR, Magnus P, Samuelsen SO, Tambs K. No evidence for effects of family environment on asthma. A retrospective study of Norwegian twins. *Am J Respir Crit Care Med.* 1997;156(1):43-49.

142. Laitinen T, Rasanen M, Kaprio J, Koskenvuo M, Laitinen LA. Importance of genetic factors in adolescent asthma: a population-based twin-family study. *Am J Respir Crit Care Med.* 1998;157(4 Pt 1):1073-1078.

143. Nystad W, Roysamb E, Magnus P, Tambs K, Harris JR. A comparison of genetic and environmental variance structures for asthma, hay fever and eczema with symptoms of the same diseases: a study of Norwegian twins. *International Journal of Epidemiology.* 2005;34(6):1302-1309.

144. Willemsen G, van Beijsterveldt TC, van Baal CG, Postma D, Boomsma DI. Heritability of self-reported asthma and allergy: a study in adult Dutch twins, siblings and parents. *Twin Res Hum Genet.* 2008;11(2):132-142.

145. Thomsen SF, van der Sluis S, Kyvik KO, Skytthe A, Backer V. Estimates of asthma heritability in a large twin sample. *Clin Exp Allergy.* 2010;40(7):1054-1061.

146. Moffatt MF, Gut IG, Demenais F, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med.* 2010;363(13):1211-1221.

147. Torgerson DG, Ampleford EJ, Chiu GY, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet.* 2011;43(9):887-892.

148. Ferreira MA, Matheson MC, Duffy DL, et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet.* 2011;378(9795):1006-1014.

149. Wan YI, Shrine NR, Soler Artigas M, et al. Genome-wide association study to identify genetic determinants of severe asthma. *Thorax.* 2012;67(9):762-768.

150. Mathias RA, Grant AV, Rafaels N, et al. A genome-wide association study on African-ancestry populations for asthma. *J Allergy Clin Immunol.* 2010;125(2):336-346.e334.

151. Verlaan DJ, Berlivet S, Hunninghake GM, et al. Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet.* 2009;85(3):377-393.

152. Fuertes E, Soderhall C, Acevedo N, et al. Associations between the 17q21 region and allergic rhinitis in 5 birth cohorts. *J Allergy Clin Immunol.* 2014.

153. Sleiman PM, Flory J, Imielinski M, et al. Variants of DENND1B associated with asthma in children. *N Engl J Med.* 2010;362(1):36-44.

154. DeWan AT, Triche EW, Xu X, et al. PDE11A associations with asthma: results of a genome-wide association scan. *J Allergy Clin Immunol.* 2010;126(4):871-873.e879.

155. Noguchi E, Sakamoto H, Hirota T, et al. Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet.* 2011;7(7):e1002170.

156. Hirota T, Takahashi A, Kubo M, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet.* 2011;43(9):893-896.

157. Schauberger EM, Ewart SL, Arshad SH, et al. Identification of ATPAF1 as a novel candidate gene for asthma in children. *J Allergy Clin Immunol.* 2011;128(4):753-760.e711.

158. Ricci G, Astolfi A, Remondini D, et al. Pooled genome-wide analysis to identify novel risk loci for pediatric allergic asthma. *PLoS One.* 2011;6(2):e16912.

159. Forno E, Lasky-Su J, Himes B, et al. Genome-wide association study of the age of onset of childhood asthma. *J Allergy Clin Immunol.* 2012;130(1):83-90 e84.

160. Ramasamy A, Kuokkanen M, Vedantam S, et al. Genome-wide association studies of asthma in population-based cohorts confirm known and suggested loci and identify an additional association near HLA. *PLoS One.* 2012;7(9):e44008.

161. Li X, Ampleford EJ, Howard TD, et al. Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. *J Allergy Clin Immunol.* 2012;130(4):861-868.e867.

162. Lasky-Su J, Himes BE, Raby BA, et al. HLA-DQ strikes again: genome-wide association study further confirms HLA-DQ in the diagnosis of asthma among adults. *Clin Exp Allergy.* 2012;42(12):1724-1733.

163. Bonnelykke K, Sleiman P, Nielsen K, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet.* 2014;46(1):51-55.

164. Ding L, Abebe T, Beyene J, et al. Rank-based genome-wide analysis reveals the association of ryanodine receptor-2 gene variants with childhood asthma among human populations. *Hum Genomics.* 2013;7:16.

165. Galanter JM, Gignoux CR, Torgerson DG, et al. Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study. *J Allergy Clin Immunol.* 2014;134(2):295-305.

166. Ferreira MA, Matheson MC, Tang CS, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. *J Allergy Clin Immunol.* 2014;133(6):1564-1571.

167. Dijk FN, de Jongste JC, Postma DS, Koppelman GH. Genetics of onset of asthma. *Curr Opin Allergy Clin Immunol.* 2013;13(2):193-202.

168. Zhao YF, Luo YM, Xiong W, Wu XL. Genetic variation in ORMDL3 gene may contribute to the risk of asthma: a meta-analysis. *Hum Immunol.* 2014;75(9):960-967.

169. Rossnerova A, Tulupova E, Tabashidze N, et al. Factors affecting the 27K DNA methylation pattern in asthmatic and healthy children from locations with various environments. *Mutat Res.* 2013;741-742:18-26.

170. Breton CV, Siegmund KD, Joubert BR, et al. Prenatal tobacco smoke exposure is associated with childhood DNA CpG methylation. *PLoS One.* 2014;9(6):e99716.

171. Yang IV, Pedersen BS, Liu A, et al. DNA methylation and childhood asthma in the inner city. *J Allergy Clin Immunol.* 2015;136(1):69-80.

172. Murphy TM, Wong CC, Arseneault L, et al. Methylomic markers of persistent childhood asthma: a longitudinal study of asthma-discordant monozygotic twins. *Clin Epigenetics.* 2015;7:130.

173. Mukherjee N, Lockett GA, Merid SK, et al. DNA methylation and genetic polymorphisms of the Leptin gene interact to influence lung function outcomes and asthma at 18 years of age. *Int J Mol Epidemiol Genet.* 2016;7(1):1-17.

174. Somineni H, Zhang X, Biagini-Myers J, et al. Ten-eleven translocation 1 (TET1) methylation is associated with childhood asthma and traffic-related air pollution. *J Allergy Clin Immunol.* 2016;137(3):797-805 e795.

175. Fu A, Leaderer BP, Gent JF, Leaderer D, Zhu Y. An environmental epigenetic study of ADRB2 5'-UTR methylation and childhood asthma severity. *Clin Exp Allergy.* 2012;42(11):1575-1581.

176. Acevedo N, Reinius LE, Greco D, et al. Risk of childhood asthma is associated with CpG-site polymorphisms, regional DNA methylation and mRNA levels at the GSDMB/ORMDL3 locus. *Hum Mol Genet.* 2015;24(3):875-890.

177. Runyon RS, Cachola LM, Rajeshuni N, et al. Asthma Discordance in Twins Is Linked to Epigenetic Modifications of T Cells. *PLoS One.* 2012;7(11).

178. Li R, Montpetit A, Rousseau M, et al. Somatic point mutations occurring early in development: a monozygotic twin study. *J Med Genet.* 2014;51(1):28-34.

179. Abdellaoui A, Ehli EA, Hottenga JJ, et al. CNV Concordance in 1,097 MZ Twin Pairs. *Twin Res Hum Genet.* 2015;18(1):1-12.

180. Magnusson PK, Lee D, Chen X, et al. One CNV Discordance in NRXN1 Observed Upon Genome-wide Screening in 38 Pairs of Adult Healthy Monozygotic Twins. *Twin Res Hum Genet.* 2016;19(2):97-103.

181. McRae AF, Visscher PM, Montgomery GW, Martin NG. Large Autosomal Copy-Number Differences within Unselected Monozygotic Twin Pairs are Rare. *Twin Res Hum Genet.* 2015;18(1):13-18.

182. Lee HJ, Yoon SC, Ko JM, et al. Monochorionic dizygotic twins with discordant sex and confined blood chimerism. *Eur J Pediatr.* 2014;173(9):1249-1252.

183. Ananth CV, Vintzileos AM, Shen-Schwarz S, Smulian JC, Lai YL. Standards of birth weight in twin gestations stratified by placental chorionicity. *Obstet Gynecol.* 1998;91(6):917-924.

184. Hanley ML, Ananth CV, Shen-Schwarz S, Smulian JC, Lai YL, Vintzileos AM. Placental cord insertion and birth weight discordancy in twin gestations. *Obstet Gynecol.* 2002;99(3):477-482.

185. Loos RJ, Derom C, Derom R, Vlietinck R. Determinants of birthweight and intrauterine growth in liveborn twins. *Paediatr Perinat Epidemiol.* 2005;19 Suppl 1:15-22.

186. Morley R. Can we generalise from findings in twins? *Paediatr Perinat Epidemiol.* 2005;19 Suppl 1:54-59.

187. Posthuma D, Beem AL, de Geus EJC, et al. Theory and practice in quantitative genetics. *Twin Research.* 2003;6(5):361-376.

188. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet.* 2008;9(4):255-266.

189. Harju M, Keski-Nisula L, Georgiadis L, Heinonen S. Parental smoking and cessation during pregnancy and the risk of childhood asthma. *BMC Public Health.* 2016;16:428.

190. Vardavas CI, Hohmann C, Patelarou E, et al. The independent role of prenatal and postnatal exposure to active and passive smoking on the development of early wheeze in children. *Eur Respir J.* 2016;48(1):115-124.

191. Bergen SE, Gardner CO, Kendler KS. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Res Hum Genet.* 2007;10(3):423-433.

192. Socialstyrelsen. Statistics and Data. http://www.socialstyrelsen.se/statistics. Accessed 2016-07-28, 2016.

193. Asberg KH. [The Neurosedyn disaster became a starting shot. The drug approval process was changed and new era of patient safety was initiated]. *Lakartidningen.* 2011;108(48):2486-2487.

194. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol.* 2009;24(11):659-667.

195. Magnusson PK, Almqvist C, Rahman I, et al. The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res Hum Genet.* 2013;16(1):317-329.

196. Anckarsater H, Lundstrom S, Kollberg L, et al. The Child and Adolescent Twin Study in Sweden (CATSS). *Twin Research and Human Genetics.* 2011;14(6):495-508.

197. Hannelius U, Gherman L, Makela VV, et al. Large-scale zygosity testing using single nucleotide polymorphisms. *Twin Res Hum Genet.* 2007;10(4):604-625.

198. Larson T, Anckarsater H, Gillberg C, et al. The autism--tics, AD/HD and other comorbidities inventory (A-TAC): further validation of a telephone interview for epidemiological research. *BMC Psychiatry.* 2010;10:1.

199. Ullemar V, Lundholm C, Ortqvist AK, et al. Predictors of adolescents' consent to use health records for research and results from data collection in a Swedish twin cohort. *Twin Res Hum Genet.* 2015;18(3):256-265.

200. Socialstyrelsen. *Vägledning för barnhälsovården.* 2014.

201. Socialstyrelsen. *Vägledning för elevhälsan.* 2014.

202. Almqvist C, Ortqvist AK, Ullemar V, Lundholm C, Lichtenstein P, Magnusson PK. Cohort Profile: Swedish Twin Study on Prediction and Prevention of Asthma (STOPPA). *Twin Res Hum Genet.* 2015;18(3):273-280.

203. National Board of Health and Welfare Centre for Epidemiology. *The Swedish Medical Birth Register - A summary of content and quality*. 2003; http://www.socialstyrelsen.se/publikationer2003/2003-112-3 Accessed 2016-08-20

204. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health.* 2011;11:450.

205. Wettermark B, Hammar N, Fored CM, et al. The new Swedish Prescribed Drug Register--opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf.* 2007;16(7):726-735.

206. *WHO Collaborating Centre for Drug Statistics Methodology,*

Guidelines for ATC classification and DDD assignment. 16 ed2013.

207. Ortqvist AK, Lundholm C, Wettermark B, Ludvigsson JF, Ye W, Almqvist C. Validation of asthma and eczema in population-based Swedish drug and patient registers. *Pharmacoepidemiol Drug Saf.* 2013;22(8):850-860.

208. Almqvist C, Cnattingius S, Lichtenstein P, Lundholm C. The impact of birth mode of delivery on childhood asthma and allergic diseases--a sibling study. *Clin Exp Allergy.* 2012;42(9):1369-1376.

209. Andolf E, Lilja H, Hildingsson I, et al. Socialstyrelsen. *Nationella Medicinska Indikationer Rapport 2011:9 Indikation för kejsarsnitt på moderns önskan.* 2011.

210. Cole TJ, Donaldson MD, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. *Int J Epidemiol.* 2010;39(6):1558-1566.

211. Ekholm L. *Rikshandboken Barnhälsovård: Vid vilka åldrar och hur mäter BVC barn*. 2016; http://www.rikshandboken-bhv.se/Texter/Barns-tillvaxt-0-6-ar/Vid-vilka-aldrar-och-hur-mater-BVC-barn/. Accessed 2016-08-20.

212. Levin AM, Mathias RA, Huang L, et al. A meta-analysis of genome-wide association studies for serum total IgE in diverse study populations. *J Allergy Clin Immunol.* 2013;131(4):1176-1184.

213. Buhule OD, Minster RL, Hawley NL, et al. Stratified randomization controls better for batch effects in 450K methylation analysis: a cautionary tale. *Front Genet.* 2014;5:354.

214. Bell JT, Spector TD. A twin approach to unraveling epigenetics. *Trends Genet.* 2011;27(3):116-125.

215. Fahy J. Eosinophilic and Neutrophilic Inflammation in Asthma. *Proceedings of the American Thoracic Society.* 2009;6(3):256-259.

216. VanderWeele TJ, Hernan MA, Robins JM. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology.* 2008;19(5):720-728.

217. VanderWeele T, Shpitser I. On the definition of a confounder. *Annals of Statistics.* 2013;41(1):196-220.

218. Hernán M, Robins J. Causal diagrams and conditional independence. *Causal Inference*. Boca Raton: Chapman & Hall/CRC; 2016:73-77.

219. Pearl J. Causal diagrams for empirical research. *Biometrika.* 1995;82:669-710.

220. Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics.* 2000;56(2):645-646.

221. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika.* 1982;69(1):239-241.

222. Pizzi C, Cole TJ, Richiardi L, dos-Santos-Silva I, Corvalan C, De Stavola B. Prenatal influences on size, velocity and tempo of infant growth: findings from three contemporary cohorts. *PLoS One.* 2014;9(2):e90291.

223. Tate AR, Dezateux C, Cole TJ, Millennium Cohort Study Child Health G. Is infant growth changing? *Int J Obes (Lond).* 2006;30(7):1094-1096.

224. Verweij KJ, Mosing MA, Zietsch BP, Medland SE. Estimating heritability from twin studies. *Methods Mol Biol.* 2012;850:151-170.

225. Purcell S. Statistical Methods in Behavioral Geneticcs. In: Plomin R, DeFries J, Knopik V, Neiderhiser J, eds. *Behavioral Genetics*. 6 ed. New York: Worth Publishers; 2013:357-402.

226. Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry.* 2016;73(4):354-361.

227. Boker S, Neale M, Maes H, et al. OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika.* 2011;76(2):306-317.

228. Almqvist C, Worm M, Leynaert B. Impact of gender on asthma in childhood and adolescence: a GA2LEN review. *Allergy.* Vol 63. Denmark2008:47-57.

229. Eley TC. Sex-Limitation Models. *Encyclopedia of Statistics in Behavioral Science*: John Wiley & Sons, Ltd; 2005.

230. Meier MH, Slutske WS, Heath AC, Martin NG. Sex differences in the genetic and environmental influences on childhood conduct disorder and adult antisocial behavior. *J Abnorm Psychol.* 2011;120(2):377-388.

231. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *Bmc Genomics.* 2013;14.

232. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3:Article3.

233. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods.* 2014;11(11):1138-1140.

234. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* 2010;11:587.

235. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nature Reviews Genetics.* 2002;3(11):872-882.

236. WMA. World Medical Association Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. World Medical Association; 1964.

237. Utbildningsdepartementet. Lag (2003:460) om etikprövning av forskning som avser människor. 2003.

238. Ullemar V, Lundholm C, Almqvist C. Twins' risk of childhood asthma mediated by gestational age and birthweight. *Clin Exp Allergy.* 2015;45(8):1328-1336.

239. Braback L, Hedberg A. Perinatal risk factors for atopic disease in conscripts. *Clinical and Experimental Allergy.* 1998;28(8):936-942.

240. Strachan DP, Moran SE, McInneny K, Smalls M. Reduced risk of hospital admission for childhood asthma among Scottish twins: record linkage study. *British Medical Journal.* 2000;321(7263):732-733.

241. Aspberg S, Dahlquist G, Kahan T, Kallen B. Is neonatal phototherapy associated with an increased risk for hospitalized childhood bronchial asthma? *Pediatr Allergy Immunol.* 2007;18(4):313-319.

242. Huovinen E, Kaprio J. Twins and asthma - Difference in admission rates may be due to other factors. *British Medical Journal.* 2001;322(7285):556-556.

243. Sicignano N, Beydoun HA, Russell H, Jones H, Oehninger S. A descriptive study of asthma in young adults conceived by IVF. *Reproductive Biomedicine Online.* 2010;21(6):812-818.

244. Thavagnanam S, Fleming J, Bromley A, Shields MD, Cardwell CR. A meta-analysis of the association between Caesarean section and childhood asthma. *Clin Exp Allergy.* 2008;38(4):629-633.

245. Adams M, Doull I. Birth by caesarean section and asthma. *Clin Exp Allergy.* Vol 38. England2008:554-556.

246. Bager P, Wohlfahrt J, Westergaard T. Caesarean delivery and risk of atopy and allergic disease: meta-analyses. *Clin Exp Allergy.* 2008;38(4):634-642.

247. Tollanes MC, Moster D, Daltveit AK, Irgens LM. Cesarean section and risk of severe childhood asthma: a population-based cohort study. *J Pediatr.* 2008;153(1):112-116.

248. Roduit C, Scholtens S, de Jongste JC, et al. Asthma at 8 years of age in children born by caesarean section. *Thorax.* 2009;64(2):107-113.

249. Hemminki K, Li X. Cancer risks in twins: results from the Swedish family-cancer database. *Int J Cancer.* 2002;99(6):873-878.

250. Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational age at birth. *Am J Epidemiol.* 2011;174(9):1062-1068.

251. Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight "paradox" uncovered? *Am J Epidemiol.* 2006;164(11):1115-1120.

252. Cheung YB, Yip P, Karlberg J. Mortality of twins and singletons by gestational age: a varying-coefficient approach. *Am J Epidemiol.* 2000;152(12):1107-1116.

253. Kirkby S, Genen L, Turenne W, Dysart K. Outcomes and Milestone Achievement Differences for Very Low-Birth-Weight Multiples Compared with Singleton Infants. *American Journal of Perinatology.* 2010;27(6):439-444.

254. Baker ER, Beach ML, Craigo SD, HarveyWilkes KB, Dalton ME. A comparison of neonatal outcomes of age-matched, growth-restricted twins and growth-restricted singletons. *American Journal of Perinatology.* 1997;14(8):499-502.

255. Friedman SA, Schiff E, Kao L, Kuint J, Sibai BM. Do twins mature earlier than singletons? Results from a matched cohort study. *American Journal of Obstetrics and Gynecology.* 1997;176(6):1193-1196.

256. Ballabh P, Kumari J, AlKouatly HB, et al. Neonatal outcome of triplet versus twin and singleton pregnancies: a matched case control study. *European Journal of Obstetrics Gynecology and Reproductive Biology.* 2003;107(1):28-36.

257. Garite TJ, Clark RH, Elliott JP, Thorp JA, Pediat Obstet Perinatal Res G. Twins and triplets: The effect of plurality and growth on neonatal outcome compared with singleton infants. *American Journal of Obstetrics and Gynecology.* 2004;191(3):700-707.

258. Luo ZC, Simonet F, An N, Bao FY, Audibert F, Fraser WD. Effect on neonatal outcomes in gestational hypertension in twin compared with singleton pregnancies. *Obstetrics and Gynecology.* 2006;108(5):1138-1144.

259. Luo ZC, Simonet F, Wei SQ, Xu H, Rey E, Fraser WD. Diabetes in pregnancy may differentially affect neonatal outcomes for twins and singletons. *Diabetic Medicine.* 2011;28(9):1068-1073.

260. Cheong-See F, Schuit E, Arroyo-Manzano D, et al. Prospective risk of stillbirth and neonatal complications in twin pregnancies: systematic review and meta-analysis. *BMJ.* 2016;354:i4353.

261. Oberg S, Cnattingius S, Sandin S, Lichtenstein P, Morley R, Iliadou AN. Twinship influence on morbidity and mortality across the lifespan. *Int J Epidemiol.* 2012;41(4):1002-1009.

262. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatr.* 2013;13:59.

263. Nommsen-Rivers LA, Dewey KG. Growth of breastfed infants. *Breastfeed Med.* 2009;4 Suppl 1:S45-49.

264. Cinar ND, Alvur TM, Kose D, Nemut T. Breastfeeding twins: a qualitative study. *J Health Popul Nutr.* 2013;31(4):504-509.

265. Cai X, Wardlaw T, Brown DW. Global trends in exclusive breastfeeding. *Int Breastfeed J.* 2012;7(1):12.

266. Paepe ME. Examination of the twin placenta. *Semin Perinatol.* 2015;39(1):27-35.

267. Ullemar V, Magnusson PK, Lundholm C, et al. Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy.* 2016;71(2):230-238.

268. Ono JG, Worgall TS, Worgall S. 17q21 locus and ORMDL3: an increased risk for childhood asthma. *Pediatr Res.* 2014;75(1-2):165-170.

269. Reinius LE, Gref A, Saaf A, et al. DNA Methylation in the Neuropeptide S Receptor 1 (NPSR1) Promoter in Relation to Asthma and Environmental Factors. *PLoS One.* 2013;8(1).

270. Sircar G, Saha B, Bhattacharya SG, Saha S. Allergic asthma biomarkers using systems approaches. *Front Genet.* 2014;4:308.

271. Struhl K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* 1998;12(5):599-606.

272. Purcell S. Variance components models for gene-environment interaction in twin analysis. *Twin Res.* 2002;5(6):554-571.

273. Kahr N, Naeser V, Stensballe LG, et al. Gene-environment interaction in atopic diseases: a population-based twin study of early-life exposures. *Clin Respir J.* 2015;9(1):79-86.

274. Vaissiere T, Sawan C, Herceg Z. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutat Res.* 2008;659(1-2):40-48.

275. Yang Q, Cui J, Chazaro I, Cupples LA, Demissie S. Power and type I error rate of false discovery rate approaches in genome-wide association studies. *BMC Genet.* 2005;6 Suppl 1:S134.

276. Ronmark E, Jonsson E, Platts-Mills T, Lundback B. Different pattern of risk factors for atopic and nonatopic asthma among children--report from the Obstructive Lung Disease in Northern Sweden Study. *Allergy.* 1999;54(9):926-935.

277. Asher MI, Montefort S, Bjorksten B, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *Lancet.* Vol 368. England2006:733-743.

278. Yoo Y. Phenotypes and endotypes of severe asthma in children. *Korean J Pediatr.* 2013;56(5):191-195.

279. Nordlund B, Melen E, Schultz ES, Gronlund H, Hedlin G, Kull I. Prevalence of severe childhood asthma according to the WHO. *Respir Med.* 2014;108(8):1234-1237.

280. Lazarevic V, Gaia N, Girard M, Francois P, Schrenzel J. Comparison of DNA extraction methods in analysis of salivary bacterial communities. *PLoS One.* 2013;8(7):e67699.

281. Naumova AK, Al Tuwaijri A, Morin A, et al. Sex- and age-dependent DNA methylation at the 17q12-q21 locus associated with childhood asthma. *Hum Genet.* 2013;132(7):811-822.

282. Baccarelli A, Rusconi F, Bollati V, et al. Nasal cell DNA methylation, inflammation, lung function and wheezing in children with asthma. *Epigenomics.* 2012;4(1):91-100.

283. Ebejer JL, Medland SE, van der Werf J, et al. Contrast effects and sex influence maternal and self-report dimensional measures of attention-deficit hyperactivity disorder. *Behav Genet.* 2015;45(1):35-50.